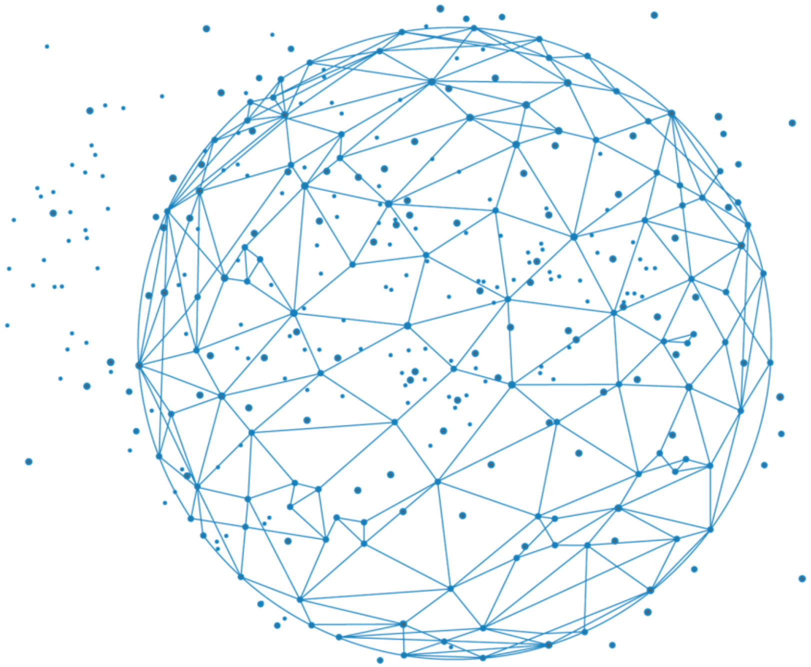
Second edition 2024 in Fréjus

ÉCOLE THÉMATIQUE ETBII BIOINFORMATIQUE INTÉGRATIVE IFB



Regularized Generalized Canonical Correlation Analysis (RGCCA)

Jimmy Vandel (Plateforme Lilloise en Biologie & Santé)
Arnaud Gloaguen (CNRGH - CEA)
Vincent Guillemot (Institut Pasteur)

DOI version final

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

4. **Unsupervised analysis with $\boldsymbol{L}$-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)

5. **Supervised analysis with RGCCA**

6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)

7. **The flexible Optimization Framework of RGCCA**

   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

4. **Unsupervised analysis with L-blocks:**
Regularized Generalized Canonical Correlation Analysis (RGCCA)

5. **Supervised analysis with RGCCA**

6. **Variable selection in RGCCA:**
Sparse Generalized Canonical Correlation Analysis (SGCCA)

7. **The flexible Optimization Framework of RGCCA**

    - The general principal
    - Extension to multi-way analysis
    - From Sequential to Global

Case Study: Major Depressive Disorder (MDD)

Data from this case study comes from Amazigh et. al 2024 **Sex-specific and multiomic integration enhance accuracy of peripheral blood biomarkers of major depressive disorder**.

a

Cohort Description

Sex	n .max	Group	mean.Age	mean.BMI	mean.HDRS	n (mRNA)	n (miRNA)	n (DNAm)
Female	55	Control	45	23.97	1.44	55 (100%)	51 (92.7%)	50 (90.9%)
	50	Patient	42	24	23.16	50 (100%)	39 (78%)	39 (78%)
Male	34	Control	45	24.02	1.18	34 (100%)	34 (100%)	31 (91.1%)
	30	Patient	43	25.51	23.47	30 (100%)	30 (100%)	28 (93.3%)
Total	**169**							

b

Data Collection

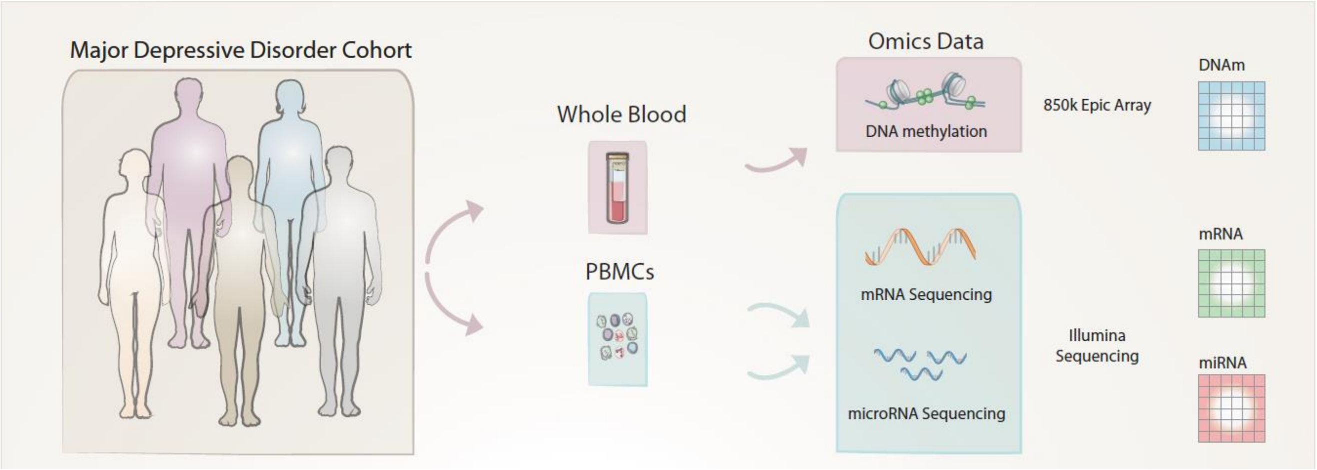


Figure taken from Amazigh Mokhtari’s PhD manuscript.

Case Study: Covariates

```
> summary(DNAm_covariates_explored_female)
  Sample_Group          BMI              BMI.bin           Age             Age.bin       Age_bin       Array                    Slide    
 control:50     Min.   :16.37    low   :64      Min.   :21.00    <20   : 0     2:21     R04C01 :19    204668820053: 3  
 mdd    :37     1st Qu.:21.42    medium:19      1st Qu.:32.00    20-30:21     3:11     R05C01 :18    204679630043: 3  
                Median :23.23    high  : 4      Median :45.00    30-40:11     4:20     R06C01 :12    204564460100: 2  
                Mean   :23.83                   Mean   :43.52    40-50:20     5:26     R07C01 :12    204564470040: 2  
                3rd Qu.:25.24                   3rd Qu.:53.50    50-60:26     6: 8     R03C01 :10    204564470092: 2  
                Max.   :39.54                   Max.   :71.00    60-70: 8     7: 1     R02C01 : 8    204564470101: 2  
                                                                 >70   : 1             (Other): 8    (Other)     :73  
      CD4                 CD8                  MO                   B                   NK                  GR        
 Min.   :0.08709    Min.   :0.02919    Min.   :0.04979    Min.   :0.00000    Min.   :0.00000    Min.   :0.3883  
 1st Qu.:0.15202    1st Qu.:0.08095    1st Qu.:0.07906    1st Qu.:0.01484    1st Qu.:0.03505    1st Qu.:0.5122  
 Median :0.19110    Median :0.10843    Median :0.08997    Median :0.02433    Median :0.05053    Median :0.5982  
 Mean   :0.18577    Mean   :0.10527    Mean   :0.09208    Mean   :0.02922    Mean   :0.05556    Mean   :0.5862  
 3rd Qu.:0.21439    3rd Qu.:0.12263    3rd Qu.:0.10495    3rd Qu.:0.03967    3rd Qu.:0.07699    3rd Qu.:0.6446  
 Max.   :0.30672    Max.   :0.19381    Max.   :0.14454    Max.   :0.13657    Max.   :0.14684    Max.   :0.7691  
```

Case Study: Covariates

Low (≤ 25), High (≥ 30).

```
> summary(DNAm_covariates_explored_female)
  Sample_Group         BMI              BMI.bin           Age            Age.bin      Age_bin        Array                  Slide
 control:50     Min.    :16.37    low    :64     Min.    :21.00    <20   : 0    2:21     R04C01 :19    204668820053:  3
 mdd    :37     1st Qu.:21.42    medium:19     1st Qu.:32.00    20-30:21    3:11     R05C01 :18    204679630043:  3
                Median :23.23    high  : 4     Median :45.00    30-40:11    4:20     R06C01 :12    204564460100:  2
                Mean   :23.83                  Mean   :43.52    40-50:20    5:26     R07C01 :12    204564470040:  2
                3rd Qu.:25.24                  3rd Qu.:53.50    50-60:26    6: 8     R03C01 :10    204564470092:  2
                Max.   :39.54                  Max.   :71.00    60-70: 8    7: 1     R02C01 : 8    204564470101:  2
                                                                >70   : 1            (Other): 8    (Other)     :73
      CD4                  CD8                  MO                   B                    NK                   GR
 Min.   :0.08709    Min.   :0.02919    Min.   :0.04979    Min.   :0.00000    Min.   :0.00000    Min.   :0.3883
 1st Qu.:0.15202    1st Qu.:0.08095    1st Qu.:0.07906    1st Qu.:0.01484    1st Qu.:0.03505    1st Qu.:0.5122
 Median :0.19110    Median :0.10843    Median :0.08997    Median :0.02433    Median :0.05053    Median :0.5982
 Mean   :0.18577    Mean   :0.10527    Mean   :0.09208    Mean   :0.02922    Mean   :0.05556    Mean   :0.5862
 3rd Qu.:0.21439    3rd Qu.:0.12263    3rd Qu.:0.10495    3rd Qu.:0.03967    3rd Qu.:0.07699    3rd Qu.:0.6446
 Max.   :0.30672    Max.   :0.19381    Max.   :0.14454    Max.   :0.13657    Max.   :0.14684    Max.   :0.7691
```

Case Study: Covariates

Low (≤ 25), High (≥ 30).

Relative to position on the DNAm chip.

```
> summary(DNAm_covariates_explored_female)
 Sample_Group       BMI            BMI.bin          Age            Age.bin     Age_bin        Array                  Slide
 control:50   Min.   :16.37   low   :64   Min.   :21.00   <20   : 0   2:21   R04C01 :19   204668820053: 3
 mdd    :37   1st Qu.:21.42   medium:19   1st Qu.:32.00   20-30:21   3:11   R05C01 :18   204679630043: 3
              Median :23.23   high  : 4   Median :45.00   30-40:11   4:20   R06C01 :12   204564460100: 2
              Mean   :23.83               Mean   :43.52   40-50:20   5:26   R07C01 :12   204564470040: 2
              3rd Qu.:25.24               3rd Qu.:53.50   50-60:26   6: 8   R03C01 :10   204564470092: 2
              Max.   :39.54               Max.   :71.00   60-70: 8   7: 1   R02C01 : 8   204564470101: 2
                                                          >70   : 1         (Other): 8   (Other)     :73
      CD4                CD8                 MO                  B                  NK                 GR
 Min.   :0.08709   Min.   :0.02919   Min.   :0.04979   Min.   :0.00000   Min.   :0.00000   Min.   :0.3883
 1st Qu.:0.15202   1st Qu.:0.08095   1st Qu.:0.07906   1st Qu.:0.01484   1st Qu.:0.03505   1st Qu.:0.5122
 Median :0.19110   Median :0.10843   Median :0.08997   Median :0.02433   Median :0.05053   Median :0.5982
 Mean   :0.18577   Mean   :0.10527   Mean   :0.09208   Mean   :0.02922   Mean   :0.05556   Mean   :0.5862
 3rd Qu.:0.21439   3rd Qu.:0.12263   3rd Qu.:0.10495   3rd Qu.:0.03967   3rd Qu.:0.07699   3rd Qu.:0.6446
 Max.   :0.30672   Max.   :0.19381   Max.   :0.14454   Max.   :0.13657   Max.   :0.14684   Max.   :0.7691
```

Case Study: Covariates

Low (≤ 25), High (≥ 30).

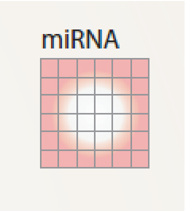
Relative to position on the DNAm chip.

```
> summary(DNAm_covariates_explored_female)
  Sample_Group        BMI            BMI.bin          Age            Age.bin     Age_bin      Array                  Slide    
 control:50     Min.   :16.37   low    :64   Min.   :21.00   <20   : 0   2:21    R04C01 :19    204668820053: 3  
 mdd    :37     1st Qu.:21.42   medium:19   1st Qu.:32.00   20-30:21   3:11    R05C01 :18    204679630043: 3  
                Median :23.23   high  : 4   Median :45.00   30-40:11   4:20    R06C01 :12    204564460100: 2  
                Mean   :23.83               Mean   :43.52   40-50:20   5:26    R07C01 :12    204564470040: 2  
                3rd Qu.:25.24               3rd Qu.:53.50   50-60:26   6: 8    R03C01 :10    204564470092: 2  
                Max.   :39.54               Max.   :71.00   60-70: 8   7: 1    R02C01 : 8    204564470101: 2  
                                                            >70   : 1          (Other): 8    (Other)     :73  
      CD4                CD8                 MO                  B                  NK                 GR        
 Min.   :0.08709   Min.   :0.02919   Min.   :0.04979   Min.   :0.00000   Min.   :0.00000   Min.   :0.3883  
 1st Qu.:0.15202   1st Qu.:0.08095   1st Qu.:0.07906   1st Qu.:0.01484   1st Qu.:0.03505   1st Qu.:0.5122  
 Median :0.19110   Median :0.10843   Median :0.08997   Median :0.02433   Median :0.05053   Median :0.5982  
 Mean   :0.18577   Mean   :0.10527   Mean   :0.09208   Mean   :0.02922   Mean   :0.05556   Mean   :0.5862  
 3rd Qu.:0.21439   3rd Qu.:0.12263   3rd Qu.:0.10495   3rd Qu.:0.03967   3rd Qu.:0.07699   3rd Qu.:0.6446  
 Max.   :0.30672   Max.   :0.19381   Max.   :0.14454   Max.   :0.13657   Max.   :0.14684   Max.   :0.7691  
```

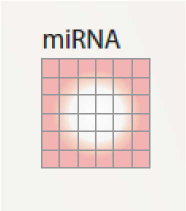
Relative to blood cell composition (T cells subsets, monocytes, B cells, NK cells and granulocytes) inferred from DNAm.

Case Study: Pre-processing

Case Study: Pre-processing

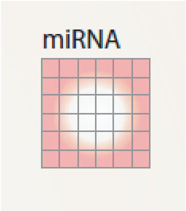


Case Study: Pre-processing



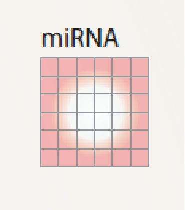
1. Remove miRNA with Nas.

Case Study: Pre-processing



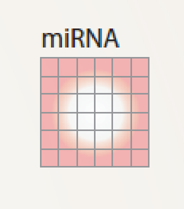
1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.

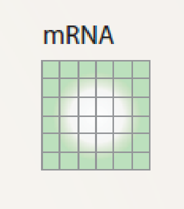
Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

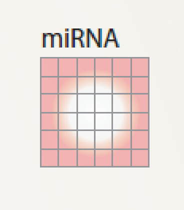
Case Study: Pre-processing



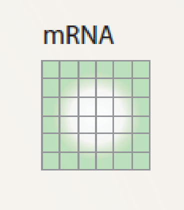


1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

Case Study: Pre-processing

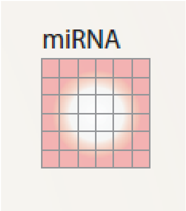


1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

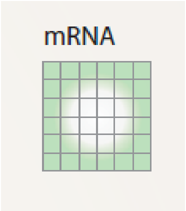


1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).

Case Study: Pre-processing

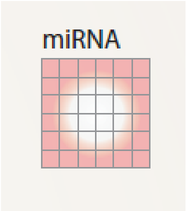


1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

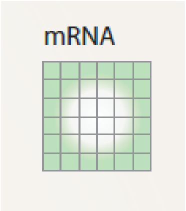


1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.

Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

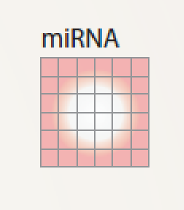


1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

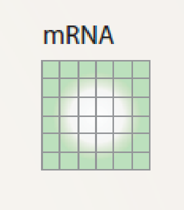
$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE — IFB — elixir FRANCE

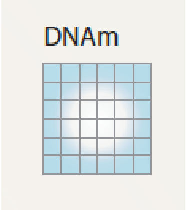
Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

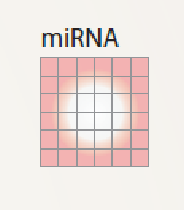


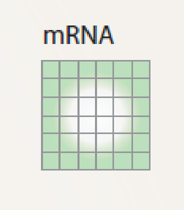
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

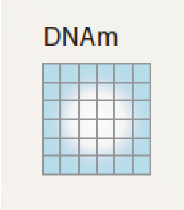


$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

Case Study: Pre-processing







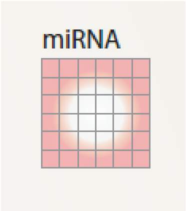
1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).

1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

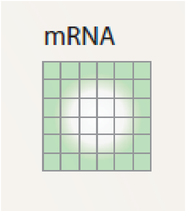
1. Normalization with Beta-MIxture Quantile (BMIQ) Normalization method (package ChAMP).

$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

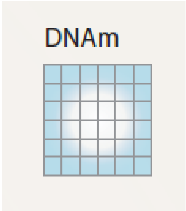
Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).



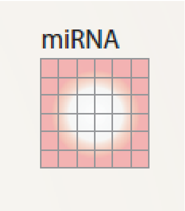
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).



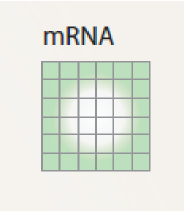
1. Normalization with Beta-MIxture Quantile (BMIQ) Normalization method (package ChAMP).
2. Remove duplicated samples.

$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

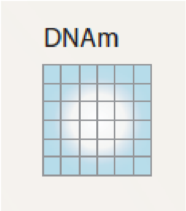
Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).



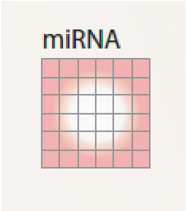
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).



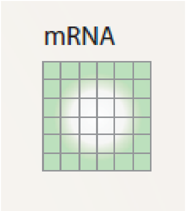
1. Normalization with Beta-MIxture Quantile (BMIQ) Normalization method (package ChAMP).
2. Remove duplicated samples.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

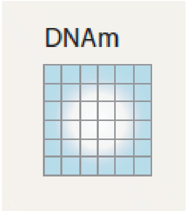
Case Study: Pre-processing



1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).



1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).



1. Normalization with Beta-MIxture Quantile (BMIQ) Normalization method (package ChAMP).
2. Remove duplicated samples.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

Finally: individuals common to ALL omics data are kept.

$MAD = median(|x_i - median(\mathbf{x})|)$, it is a robust estimation of the standard deviation.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)

5. **Supervised analysis with RGCCA**

6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)

7. **The flexible Optimization Framework of RGCCA**

   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

ChAMP's representation: Kruskal-Wallis test

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.

ChAMP’s representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$\{$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \end{cases}$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} H_0: (x_1, \dots, x_n) \text{ and } (y_1, \dots, y_m) \text{ comes from the same distribution.} \\ H_1: \text{They do not.} \end{cases}$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0: (x_1, \ldots, x_n) \text{ and } (y_1, \ldots, y_m)} \text{ comes from the same distribution.} \\ \mathrm{H_1}\text{: They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE IFB
elixir FRANCE

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \ldots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \ldots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \ldots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \ldots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

$$\Rightarrow U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

$$R_2 = (n+1) + (n+2) + \cdots + (n+m) = \frac{m((n+1)+(n+m))}{2}$$

➔ $U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1 (resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \quad \Rightarrow U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$$

$$R_2 = (n+1) + (n+2) + \cdots + (n+m) = \frac{m((n+1)+(n+m))}{2} \quad \Rightarrow U_2 = nm + \frac{m(m+1)}{2} - R_2 = 0$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H_0}: (\mathrm{x_1}, \dots, \mathrm{x_n}) \text{ and } (\mathrm{y_1}, \dots, \mathrm{y_m}) \text{ comes from the same distribution.} \\ \mathrm{H_1}: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1(resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

<u>Example:</u> "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \quad \Rightarrow U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$$

$$R_2 = (n+1) + (n+2) + \cdots + (n+m) = \frac{m((n+1)+(n+m))}{2} \quad \Rightarrow U_2 = nm + \frac{m(m+1)}{2} - R_2 = 0$$

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H}_0: (\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}) \text{ and } (\mathrm{y}_1, \dots, \mathrm{y}_\mathrm{m}) \text{ comes from the same distribution.} \\ \mathrm{H}_1: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1 (resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

Example: "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \cdots + n = \frac{n(n+1)}{2} \quad \Rightarrow U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$$

$$R_2 = (n+1) + (n+2) + \cdots + (n+m) = \frac{m((n+1)+(n+m))}{2} \quad \Rightarrow U_2 = nm + \frac{m(m+1)}{2} - R_2 = 0$$

$$\Rightarrow U = 0$$

ChAMP's representation: Kruskal-Wallis test

The Kruskal-Wallis test is a generalization of the Wilcoxon-Man-Withney test that works for two samples.
They are both **non-parametric**.

The Wilcoxon-Man-Withney proposes to test the association between a continuous (ex: age) and a discrete variable (ex: sex).

Let us consider two samples $(x_1, \dots, x_n)$ and $(y_1, \dots, y_m)$. They both represent the same continuous variable but are separated by the value of the discrete one.

$$\begin{cases} \mathrm{H}_0: (\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}) \text{ and } (\mathrm{y}_1, \dots, \mathrm{y}_\mathrm{m}) \text{ comes from the same distribution.} \\ \mathrm{H}_1: \text{They do not.} \end{cases}$$

The proposed statistic is: $U = \min(U_1, U_2) = \min\left(nm + \frac{n(n+1)}{2} - R_1, nm + \frac{m(m+1)}{2} - R_2\right)$,
where R_1 (resp. R_2) are the sum of the rank of the first (resp. second) sample when all samples are mixed and sorted.
If n and m are high enough, it is possible to show that U follows a Gaussian distribution centered in $\frac{nm+1}{2}$.

<u>Example:</u> "Perfect" association; the n first elements are in the first sample and the m next are in the second one. Then:

$$R_1 = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$$

$$R_2 = (n+1) + (n+2) + \dots + (n+m) = \frac{m((n+1)+(n+m))}{2}$$

➔ $U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm$

➔ $U_2 = nm + \frac{m(m+1)}{2} - R_2 = 0$

➔ $U = 0$

The test is likely to be rejected.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

ChAMP's representation: F-test

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0\colon \beta_1 = 0, \\ \end{cases}$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0 : \beta_1 = 0, \\ \mathrm{H}_1 : \beta_1 \neq 0 \end{cases}$

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0 : \beta_1 = 0, \\ \mathrm{H}_1 : \beta_1 \neq 0 \end{cases}$

With the statistic:

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0 : \beta_1 = 0, \\ \mathrm{H}_1 : \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{\left(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2\right)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n - 2$ degrees of freedom.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0 : \beta_1 = 0, \\ \mathrm{H}_1 : \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

If the variables are strongly linked

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} H_0 : \beta_1 = 0, \\ H_1 : \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{\left(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2\right)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n-2$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0 : \beta_1 = 0, \\ \mathrm{H}_1 : \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n-2$ ➔ The test is likely to be rejected.

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{\left(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2\right)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n - 2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n - 2$ ➔ The test is likely to be rejected.

If the variables are NOT strongly linked

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n-2$ ➔ The test is likely to be rejected.

If the variables are NOT strongly linked ➔ $RSS_0 \sim RSS_1$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n - 2$ degrees of freedom.

If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n-2$ ➔ The test is likely to be rejected.

If the variables are NOT strongly linked ➔ $RSS_0 \sim RSS_1$ ➔ $F \sim 0$

ChAMP's representation: F-test

In order to evaluate the link between two continuous variables y_i and x_i, the following model is used:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Where β_0 is the intercept, β_1 is the coefficient associated with x_i and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

The following test is used: $\begin{cases} \mathrm{H}_0: \beta_1 = 0, \\ \mathrm{H}_1: \beta_1 \neq 0 \end{cases}$

With the statistic:

$$F = \frac{(RSS_0 - RSS_1)}{1} \times \frac{(n-2)}{RSS_1} = \frac{\left(\sum_i (y_i - \beta_0)^2 - \sum_i (y_i - \beta_0 - \beta_1 x_i)^2\right)}{1} \times \frac{(n-2)}{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}$$

It is possible to show that F follows an F-distribution of 1 and $n-2$ degrees of freedom.

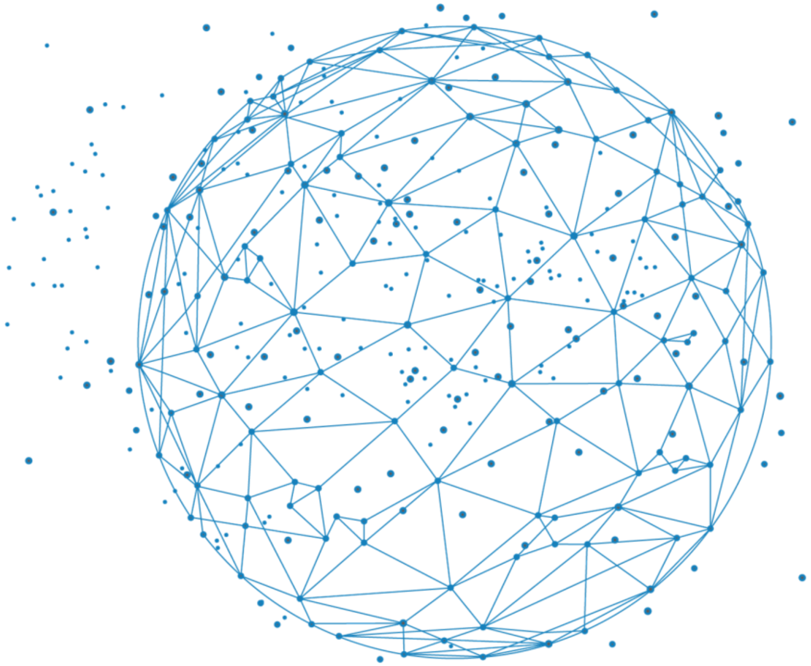
If the variables are strongly linked ➔ $RSS_0 \gg RSS_1$ ➔ $F \sim \frac{RSS_0}{RSS_1} \times (n-2) \gg n-2$ ➔ The test is likely to be rejected.

If the variables are NOT strongly linked ➔ $RSS_0 \sim RSS_1$ ➔ $F \sim 0$ ➔ The test is likely to be accepted.

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE elixir FRANCE

Now with this 2 tests, let us see what are the results of PCA on the MDD case study

➔ See section 1 on the Rmarkdown `MDD_case_study_RGCCA`



1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)

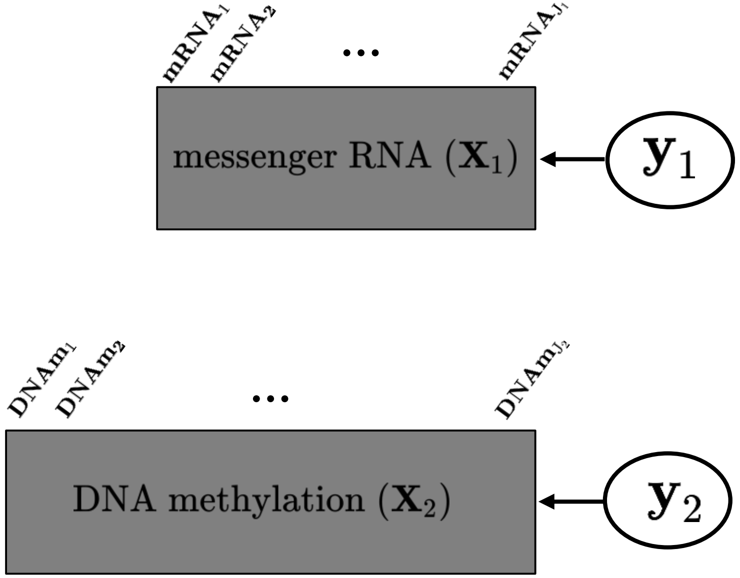
5. **Supervised analysis with RGCCA**

6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)

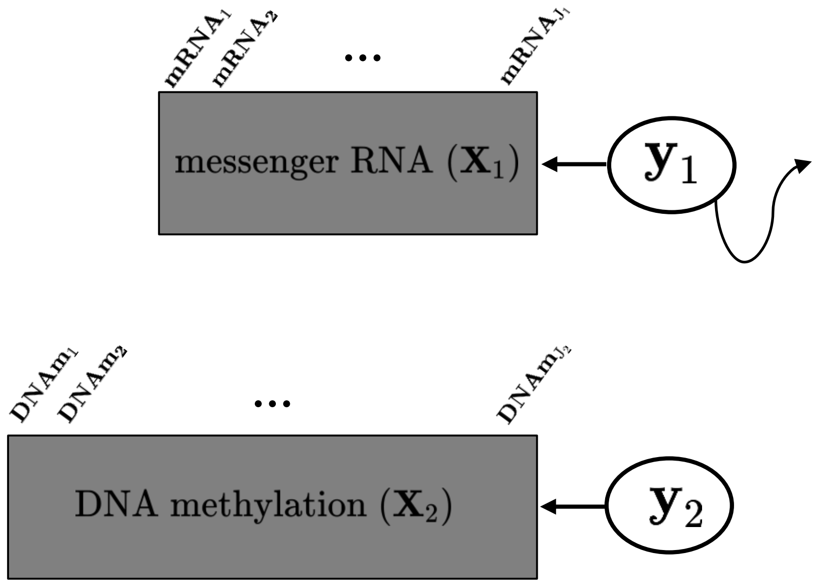
7. **The flexible Optimization Framework of RGCCA**

   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

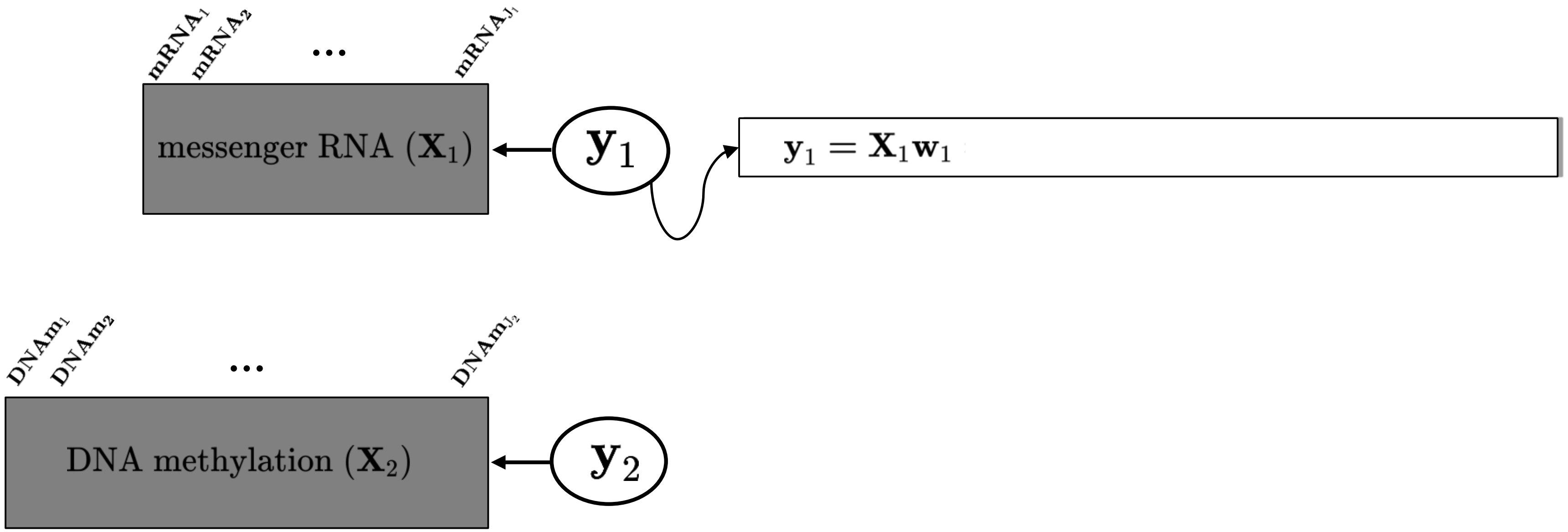
The philosophy of multiblock component methods



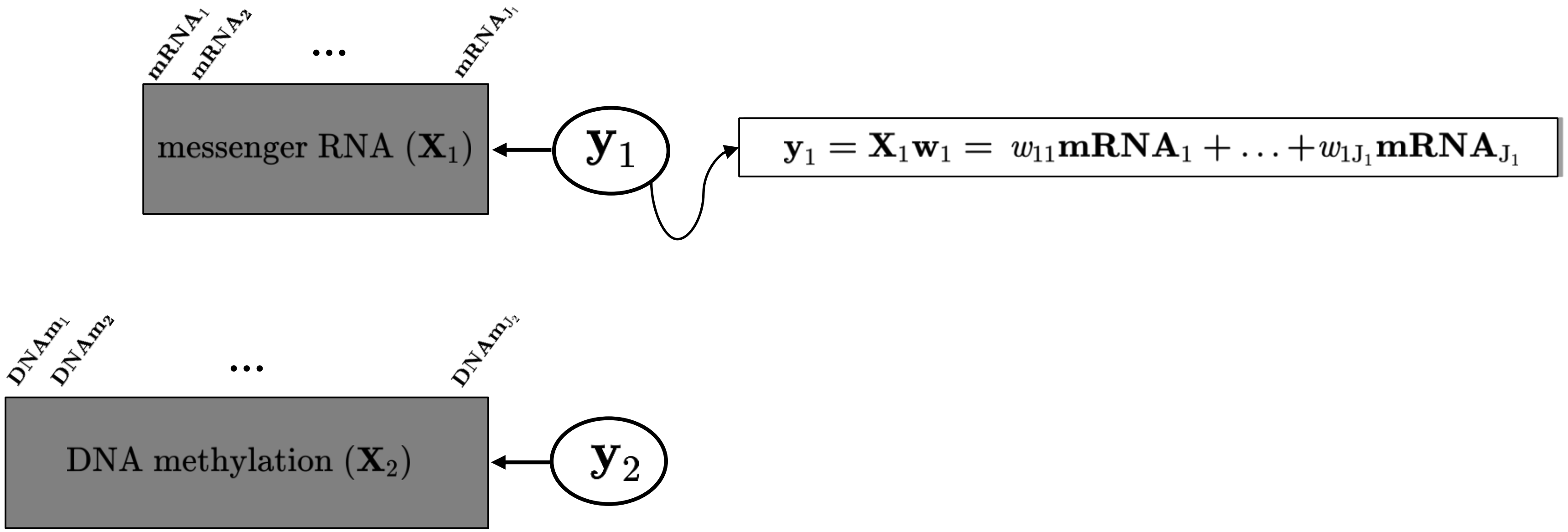
The philosophy of multiblock component methods



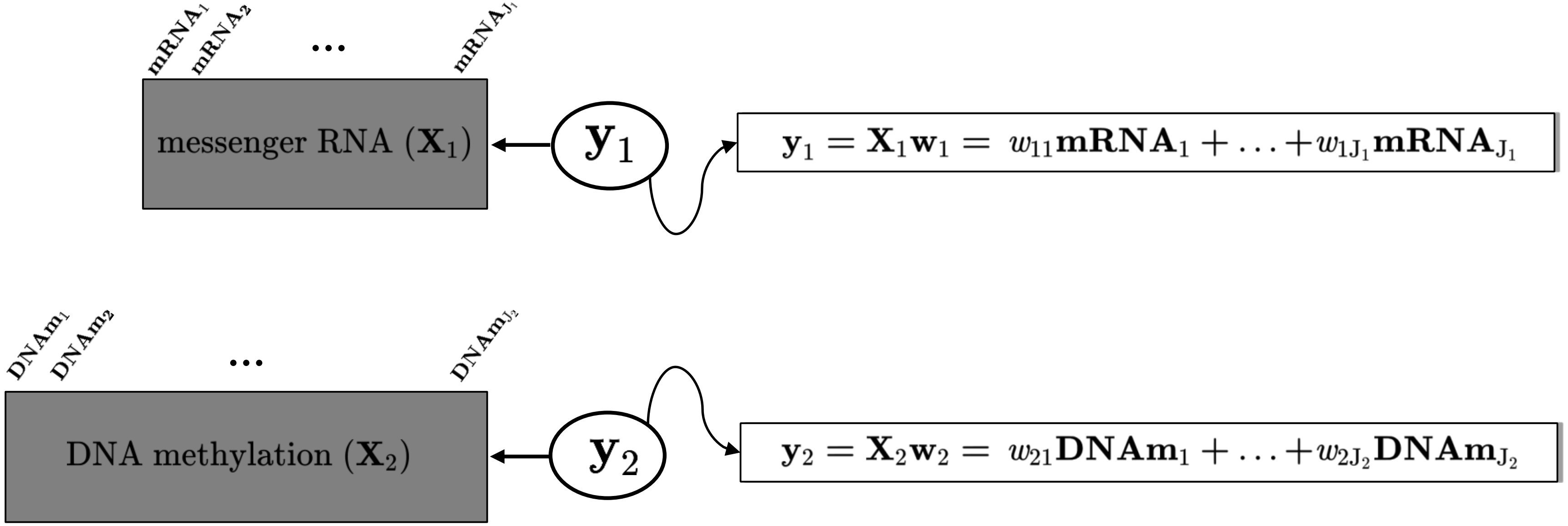
The philosophy of multiblock component methods



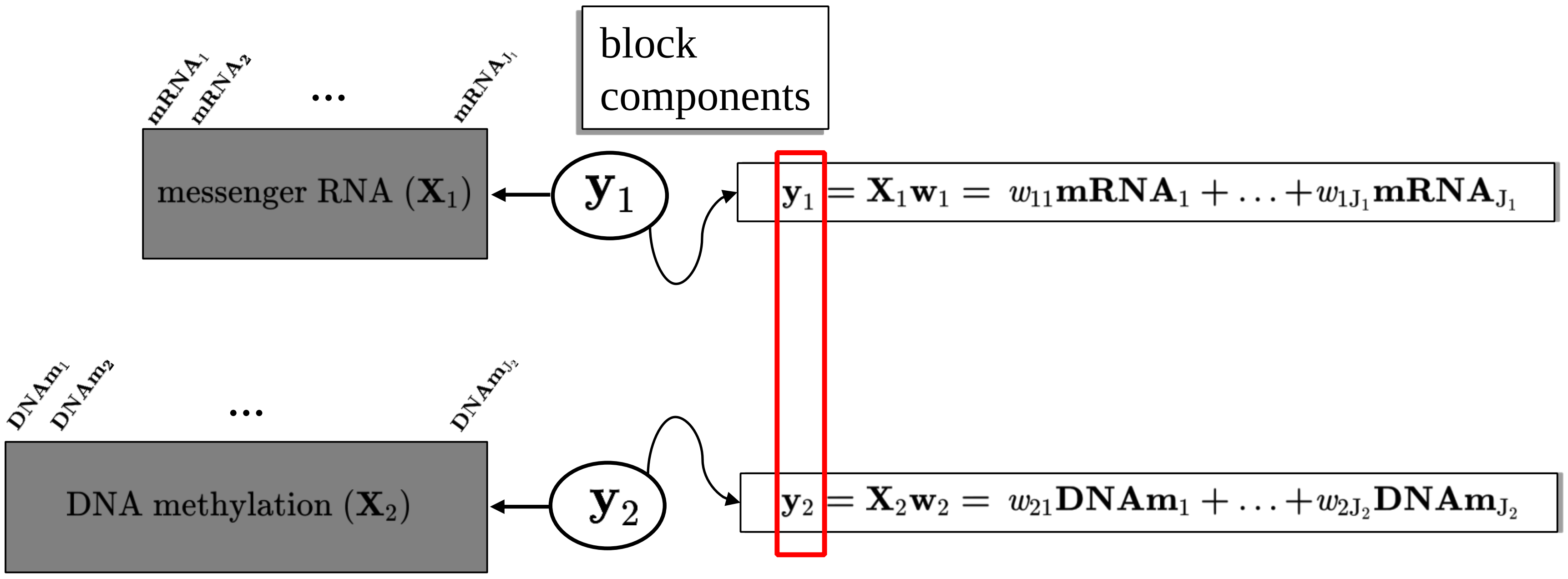
The philosophy of multiblock component methods



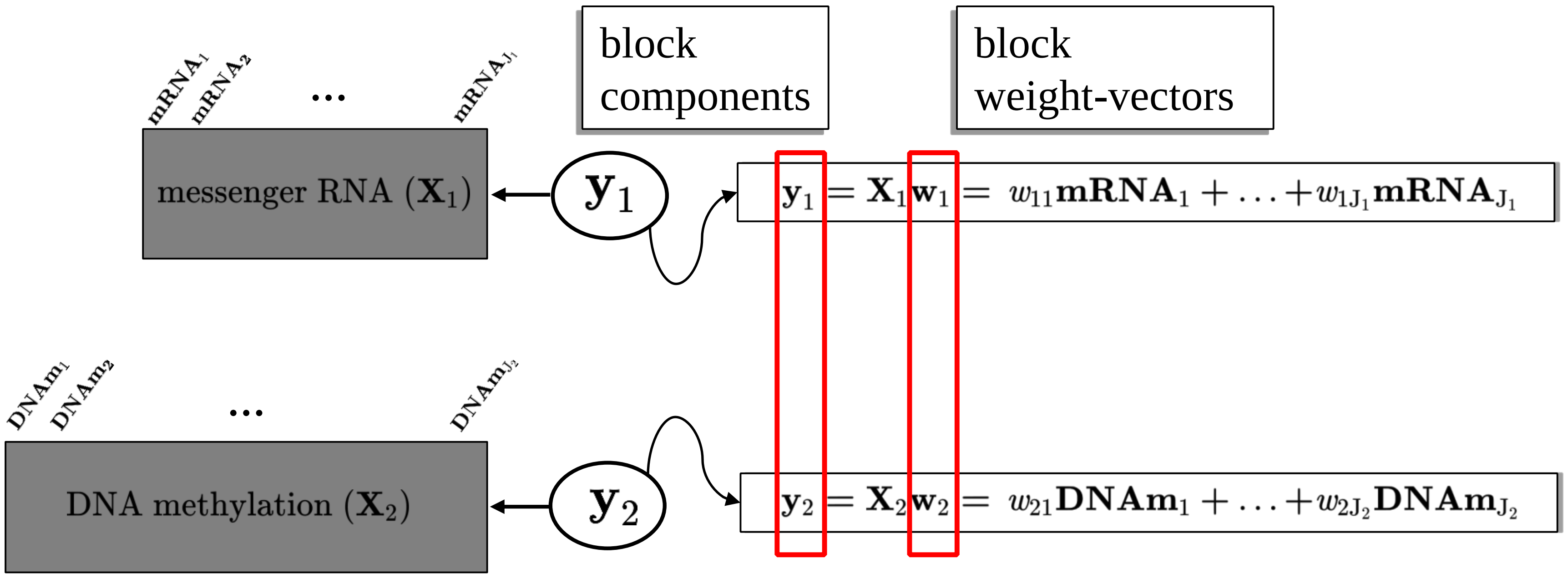
The philosophy of multiblock component methods



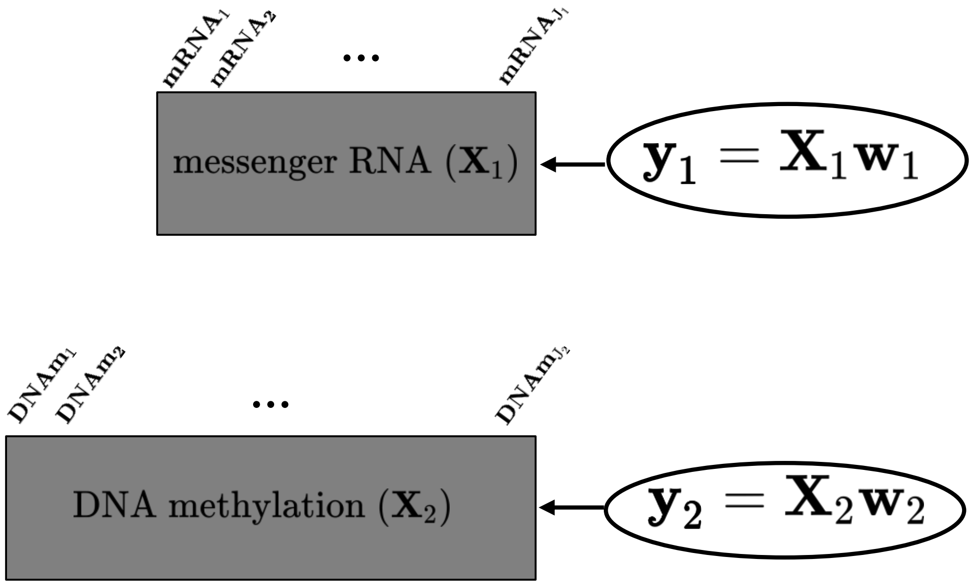
The philosophy of multiblock component methods



The philosophy of multiblock component methods



The philosophy of multiblock component methods

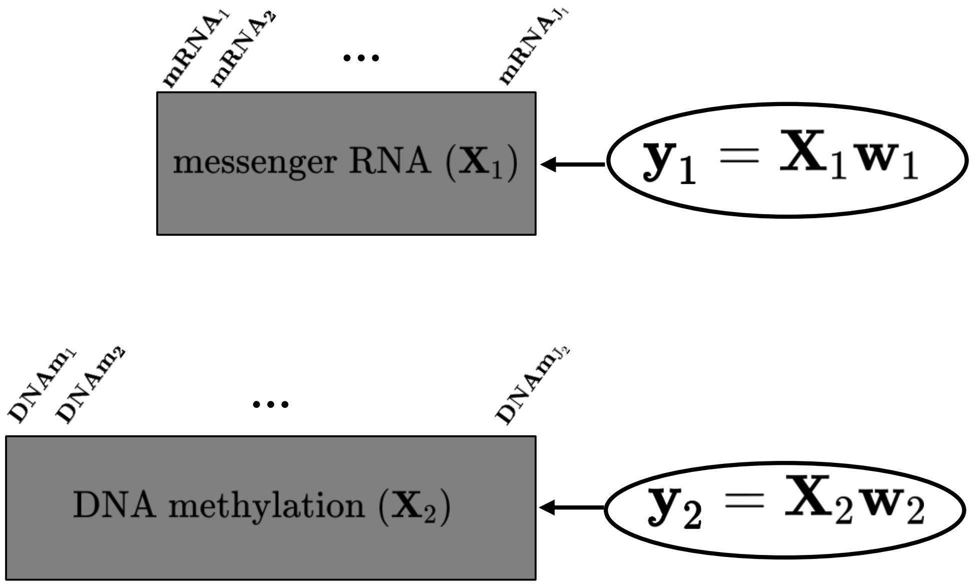


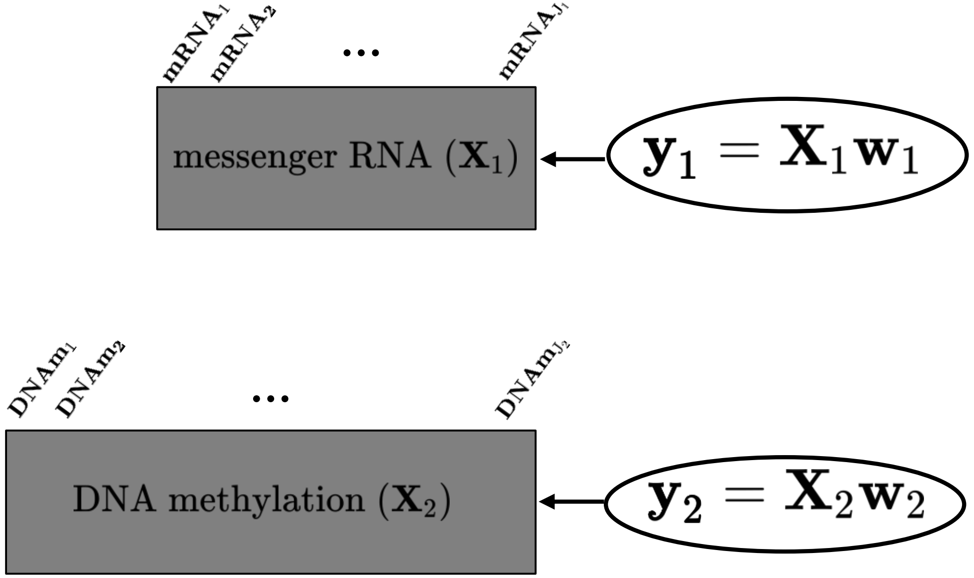
Block components should verified two properties at the same time:

1. Block components well explain their own block.

1. Block components are as correlated as possible for connected blocks.

The philosophy of multiblock component methods





Correlation based methods

Find block-weight vectors $\mathbf{w}_1, \ldots, \mathbf{w}_J$ maximizing a function of $\mathbf{\Phi} = \{\text{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\}$.

Covariance based methods

Find block-weight vectors $\mathbf{w}_1, \ldots, \mathbf{w}_J$ maximizing a function of $\mathbf{\Psi} = \{\text{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\}$.

Courtesy to Arthur Tenenhaus.

Principal Component Analysis (PCA)

$$\max_{\substack{\mathbf{w} \\ \|\mathbf{w}\|_2^2=1}} \ \mathrm{Var}(\mathbf{X}\mathbf{w})$$

From PCA to PLS/CCA

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w} \\ \|\mathbf{w}\|_2^2=1}} \ \mathrm{Var}(\mathbf{Xw})$$

From PCA to PLS/CCA

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_\mathrm{i}\|_2^2 = 1}} \mathrm{Var}(\mathbf{X}_1\mathbf{w}_1)\,\mathrm{Var}(\mathbf{X}_2\mathbf{w}_2)$$

From PCA to PLS/CCA

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_\mathrm{i}\|_2^2 = 1}} \quad \mathrm{Var}(\mathbf{X}_1\mathbf{w}_1) \qquad\qquad \mathrm{Var}(\mathbf{X}_2\mathbf{w}_2)$$

From PCA to PLS/CCA

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2 = 1}} \ \mathrm{Var}(\mathbf{X}_1\mathbf{w}_1)\,\mathrm{Cor}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)\mathrm{Var}(\mathbf{X}_2\mathbf{w}_2)$$

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

From PCA to PLS/CCA

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_\mathrm{i}\|_2^2 = 1}} \sqrt{\mathrm{Var}(\mathbf{X}_1\mathbf{w}_1)}\,\mathrm{Cor}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)\sqrt{\mathrm{Var}(\mathbf{X}_2\mathbf{w}_2)}$$

How can we "adapt" PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_\mathrm{i}\|_2^2 = 1}} \underbrace{\sqrt{\mathrm{Var}(\mathbf{X}_1\mathbf{w}_1)}\,\mathrm{Cor}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)\sqrt{\mathrm{Var}(\mathbf{X}_2\mathbf{w}_2)}}_{\mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)}$$

From PCA to PLS/CCA

<u>How can we “adapt” PCA for two-blocks analysis ?</u>

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_\mathrm{i}\|_2^2 = 1}} \quad \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

From PCA to PLS/CCA

<u>How can we "adapt" PCA for two-blocks analysis ?</u>

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \quad \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

From PCA to PLS/CCA

How can we “adapt” PCA for two-blocks analysis ?

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2) = \max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \mathrm{Cor}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

From PCA to PLS/CCA

From PCA to PLS/CCA

<u>Partial Least Squares</u> (PLS2)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2=1}} \quad \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

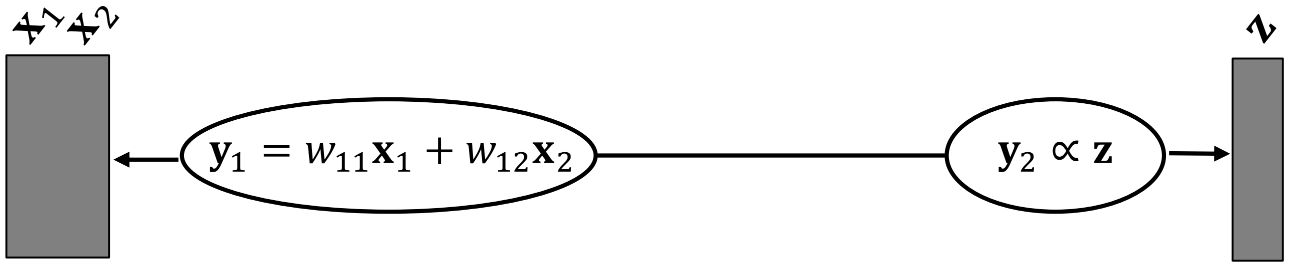
Canonical Correlation Analysis (CCA)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Partial Least Squares (PLS2)

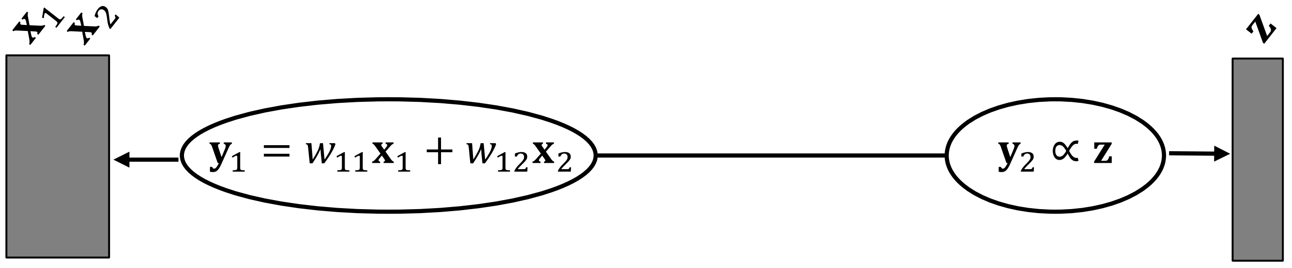
$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

PLS & CCA with a figure



PLS & CCA with a figure

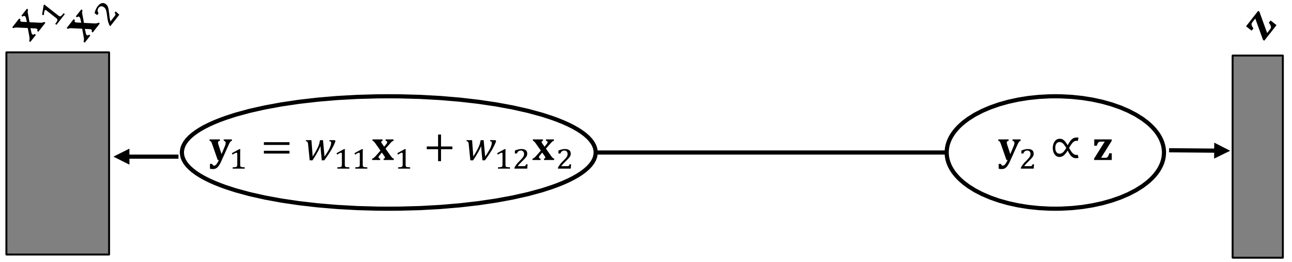
$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$



PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

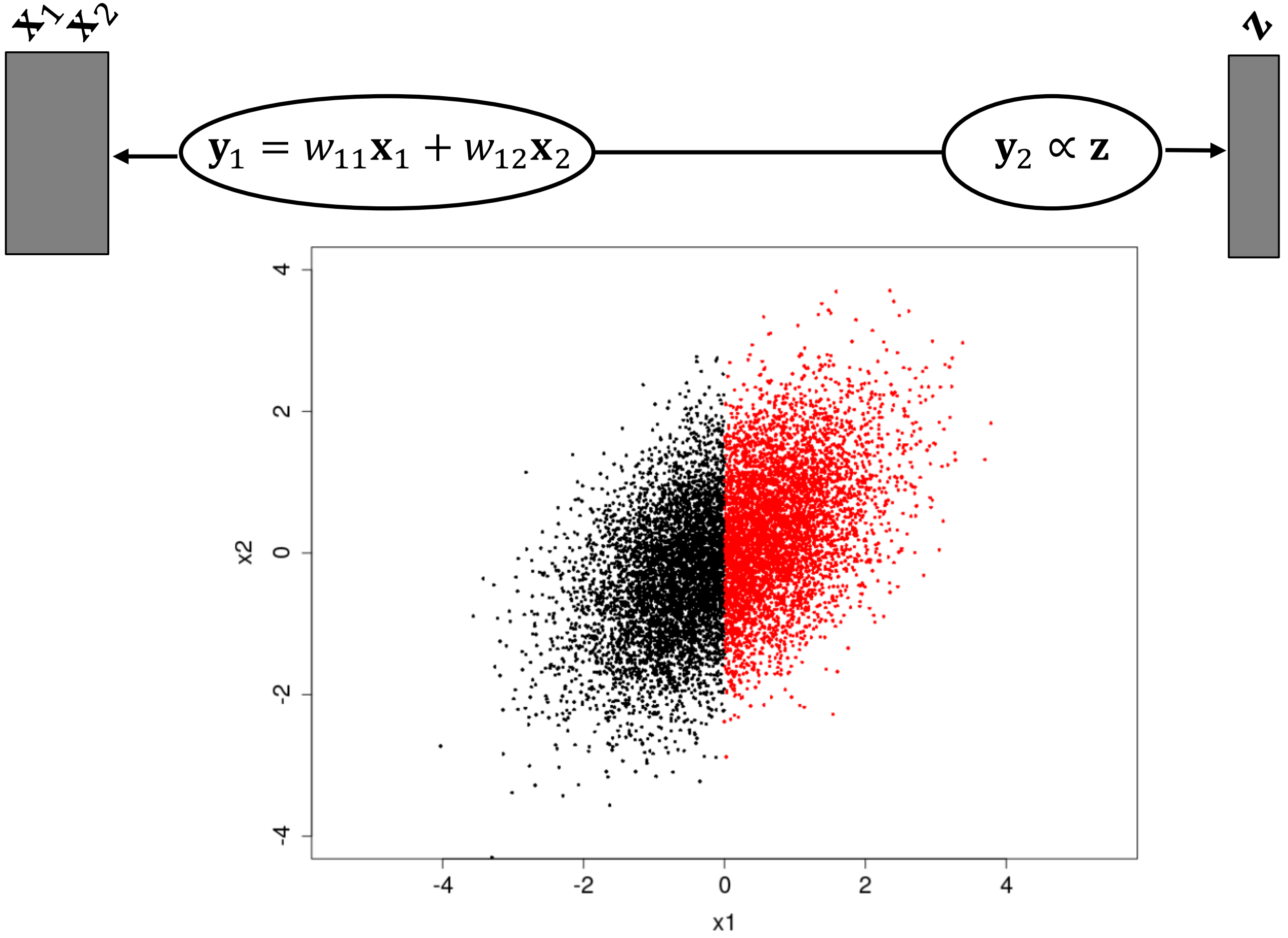
$$(\boldsymbol{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

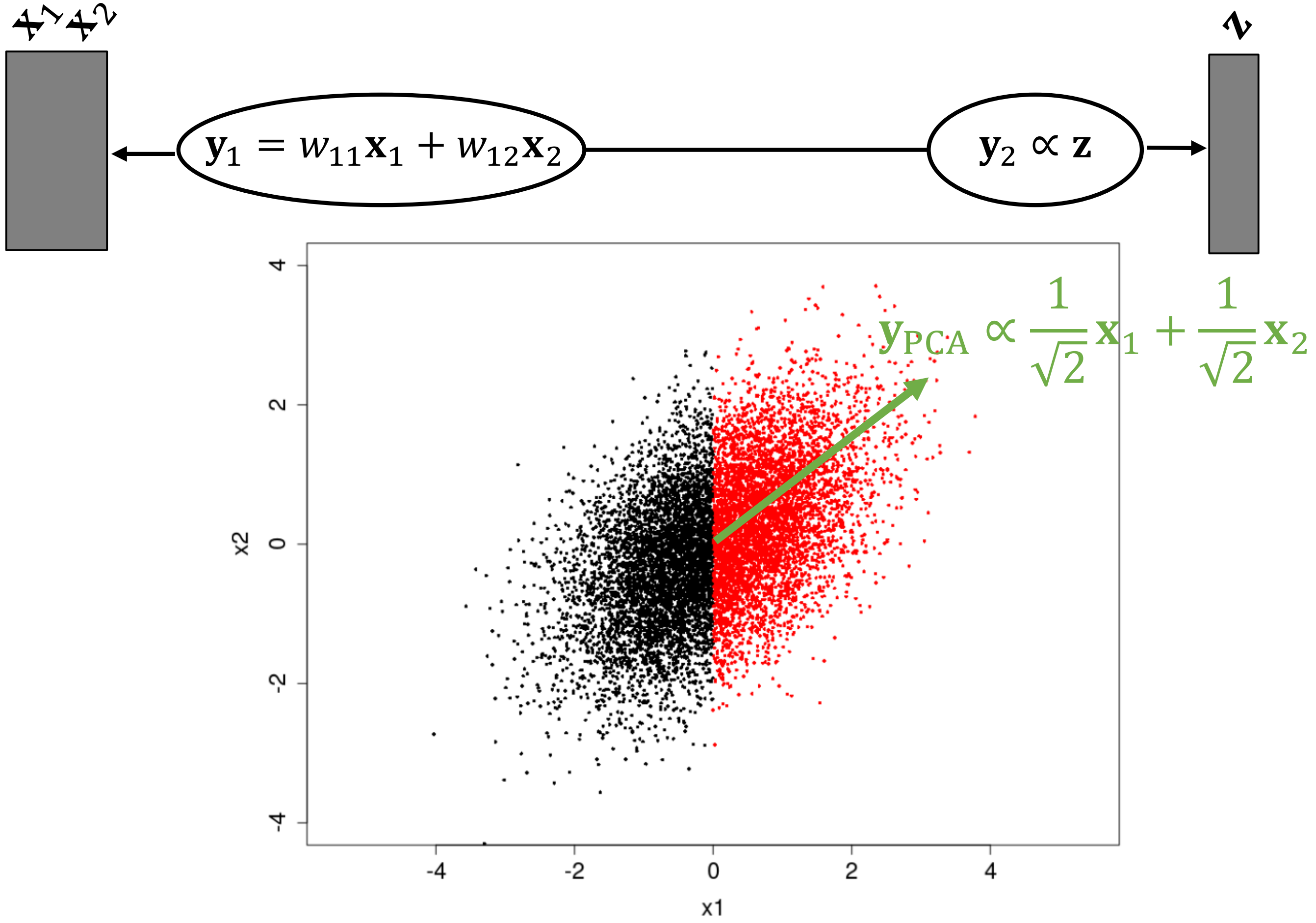
$$(\boldsymbol{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

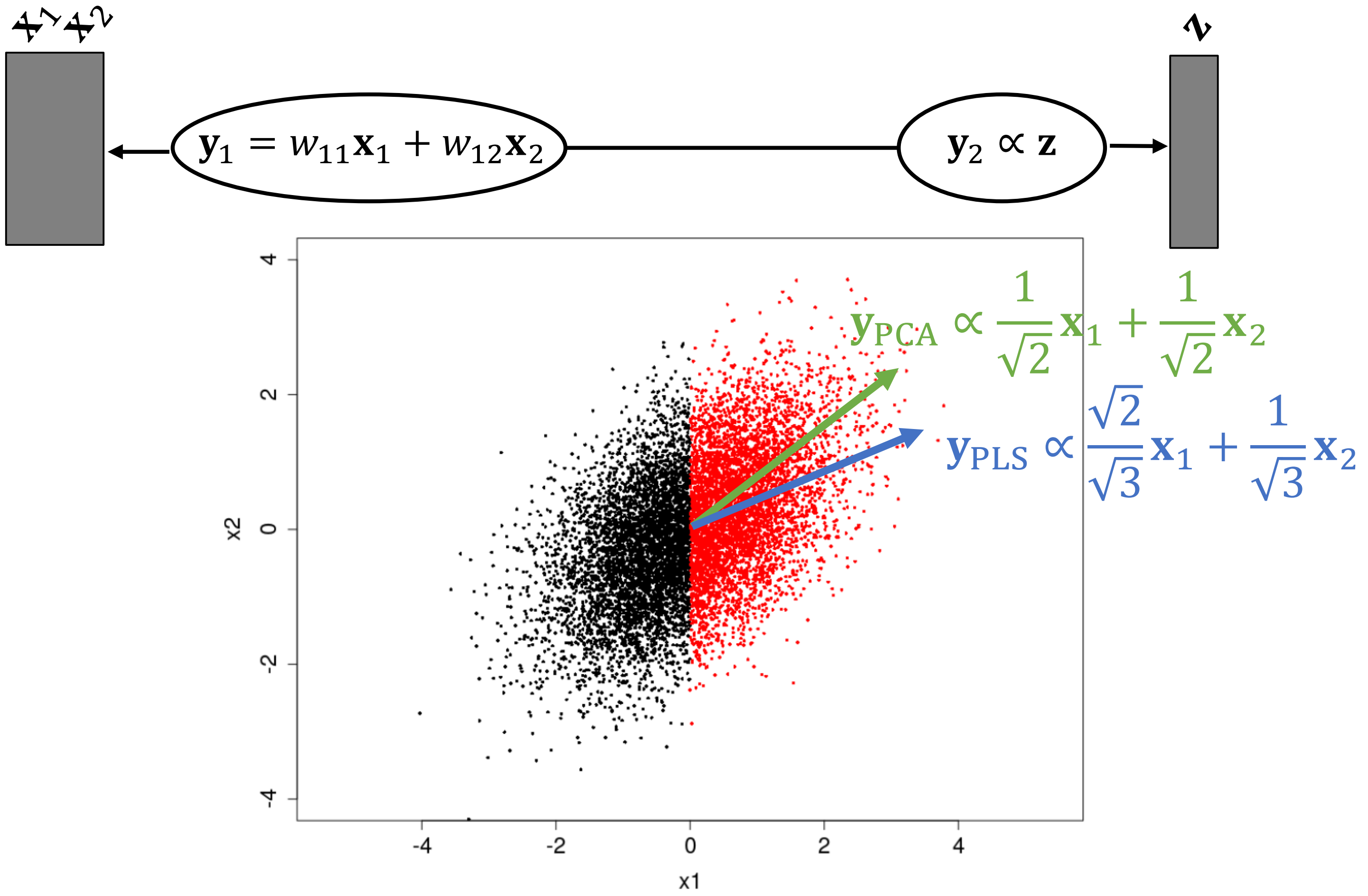
$$(\mathbf{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

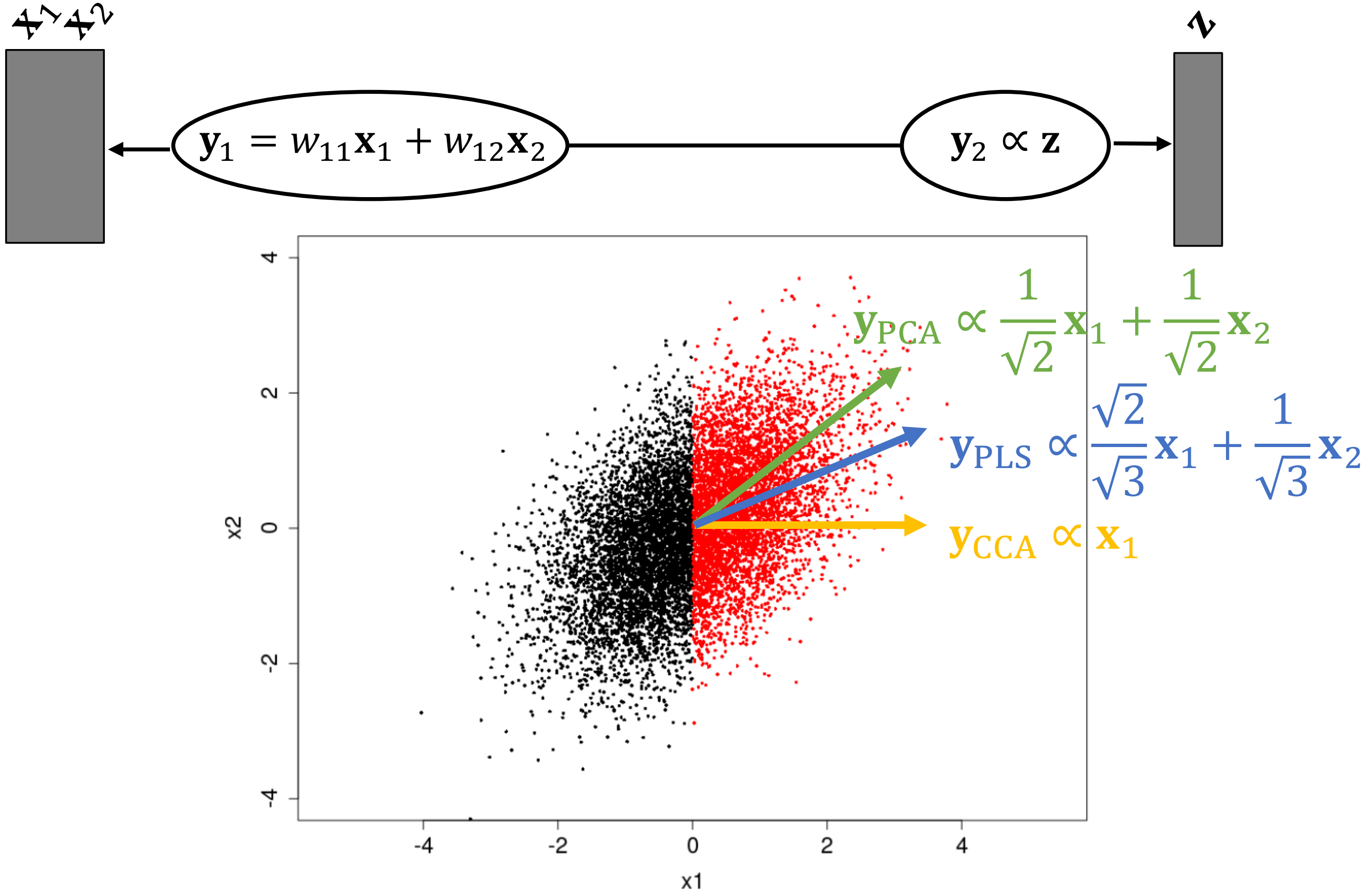
$$(\mathbf{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

$$(\boldsymbol{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



Let us see what are the results of PLS/CCA on the MDD case study

➔ See section 2.2 & 2.3 on the Rmarkdown `MDD_case_study_RGCCA`

Overfitting

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST: Subj1

TRAIN: Subj2–Subj4

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1) / TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1) / TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1)

TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Here, we are in "high-dimension" as $n < p$. The problem is ill-posed (more unknown parameters than equations).

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE elixir FRANCE

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1) / TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Here, we are in "high-dimension" as $n < p$. The problem is ill-posed (more unknown parameters than equations).

➔ It is possible to find an infinite number of solutions:

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1) / TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Here, we are in "high-dimension" as $n < p$. The problem is ill-posed (more unknown parameters than equations).

➔ It is possible to find an infinite number of solutions:

	β_1	β_2	β_3	β_4	J_{train}	J_{test}
Solution 1	43.75	0	1.375	6.25	8.4e-22	1491.891
Solution 2	-7456.25	-1000	251.375	2506.25	1.1e-19	95817179
⋮						

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1)

TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Here, we are in "high-dimension" as $n < p$. The problem is ill-posed (more unknown parameters than equations).

➔ It is possible to find an infinite number of solutions:

	β_1	β_2	β_3	β_4	J_{train}	J_{test}
Solution 1	43.75	0	1.375	6.25	8.4e-22	1491.891
Solution 2	-7456.25	-1000	251.375	2506.25	1.1e-19	95817179
⋮						

Go against the idea that age is the best explanatory variable.

Overfitting

	x_1	x_2	x_3	x_4	y
	Intercept	Age	Nb_sisters	Neighbor'weight (kg)	Subject's Height (cm)
Subj1	1	5	1	1	90
Subj2	1	10	2	50	125
Subj3	1	15	1	80	160
Subj4	1	20	2	90	180

TEST (Subj1) / TRAIN (Subj2–Subj4)

We are looking for $\beta_1, \beta_2, \beta_3$ and β_4 that minimizes $J_{TRAIN} = \sum_{i=2}^{4}(y_i - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3} - \beta_4 x_{i4})^2$.

Similarly, we can define $J_{TEST} = (y_1 - \beta_1 x_{11} - \beta_2 x_{12} - \beta_3 x_{13} - \beta_4 x_{14})^2$.

Here, we are in "high-dimension" as $n < p$. The problem is ill-posed (more unknown parameters than equations).

➔ It is possible to find an infinite number of solutions:

	β_1	β_2	β_3	β_4	J_{train}	J_{test}
Solution 1	43.75	0	1.375	6.25	8.4e-22	1491.891
Solution 2	-7456.25	-1000	251.375	2506.25	1.1e-19	95817179
⋮						

OVERFITTING

Go against the idea that age is the best explanatory variable.

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:

Cross-Validation & Regularization

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:

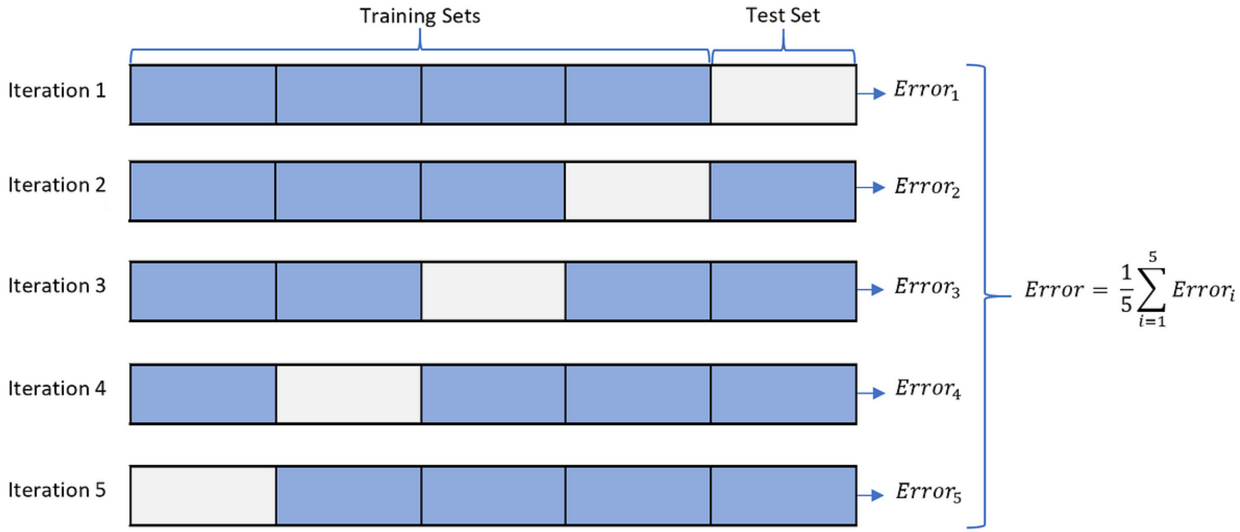
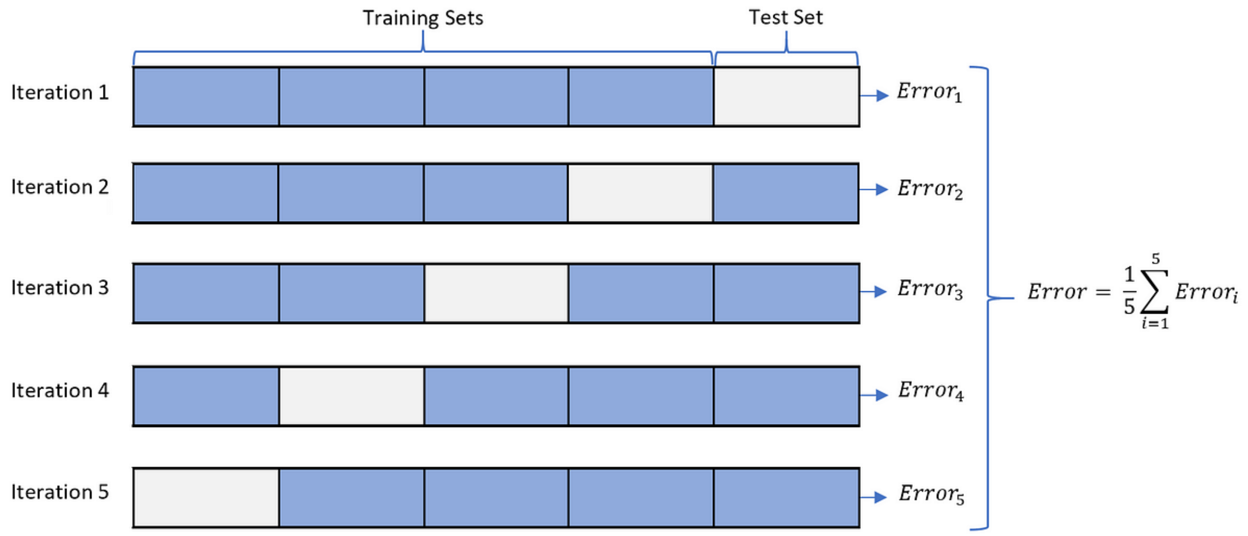


Figure extracted from https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

Cross-Validation & Regularization

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:



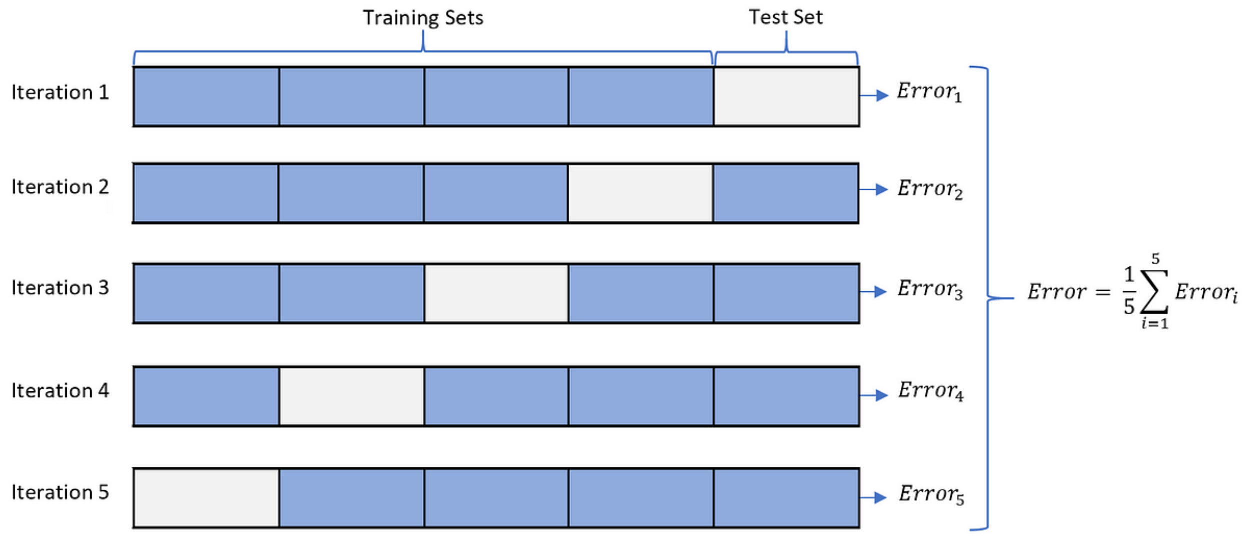
One way to avoid overfitting is by preforming regularization.

Figure extracted from https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

Cross-Validation & Regularization

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:



One way to avoid overfitting is by preforming regularization.

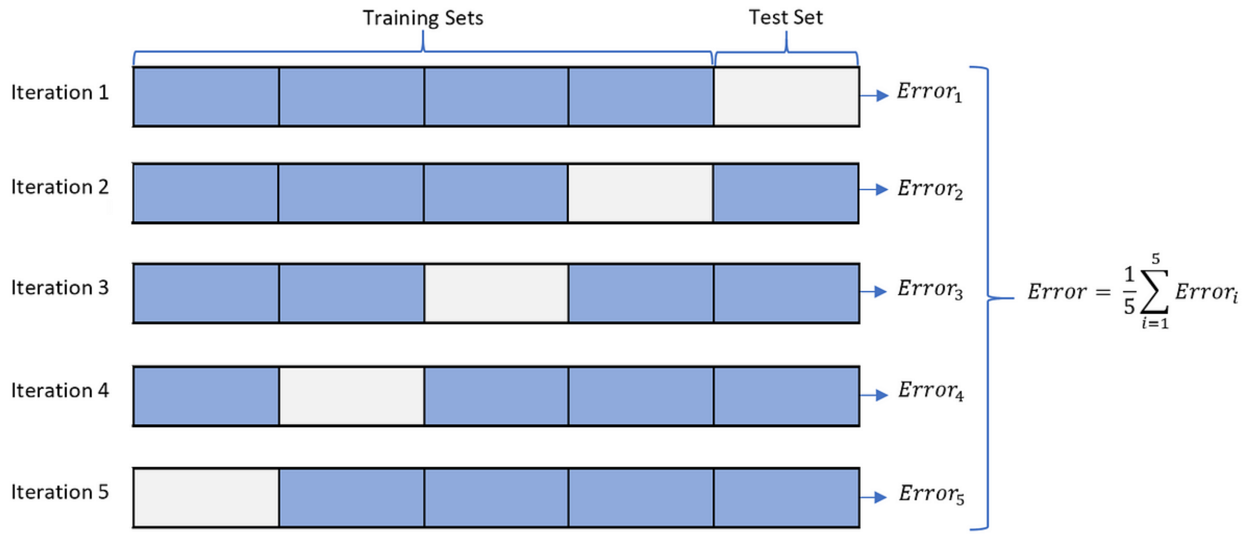
Regularization consists in adding more constraints to the model in order to reduce the space of solutions.

Figure extracted from https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

Cross-Validation & Regularization

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:



One way to avoid overfitting is by preforming regularization.

Regularization consists in adding more constraints to the model in order to reduce the space of solutions.

Multiple regularizations are available such as Ridge or LASSO regularizations.

Figure extracted from https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

Cross-Validation & Regularization

Cross-Validation allows to evaluate the generalization power of a model and realize if the model overfits or not.

A lot of sampling possibilities are available to perform Cross-Validation (CV). The most well-known is K-fold CV:

One way to avoid overfitting is by preforming regularization.

Regularization consists in adding more constraints to the model in order to reduce the space of solutions.

Multiple regularizations are available such as Ridge or LASSO regularizations.

Here, we choose to regularize the model by forcing it to have a low number of variables.

Figure extracted from https://towardsdatascience.com/cross-validation-k-fold-vs-monte-carlo-e54df2fc179b

Application on the example

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Variables considered	J_{TRAIN}	J_{TEST}
(x_1, x_2)	3.750000e+01	100
(x_1, x_3)	2.403846e+01	959.8081
(x_1, x_4)	1.512500e+03	4900
(x_1, x_2, x_3)	1.831567e-22	203.0625
(x_1, x_2, x_4)	6.464166e-24	225
(x_1, x_3, x_4)	8.664767e-22	1491.8906

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Variables considered	J_{TRAIN}	J_{TEST}
(x_1, x_2)	3.750000e+01	100
(x_1, x_3)	2.403846e+01	959.8081
(x_1, x_4)	1.512500e+03	4900
(x_1, x_2, x_3)	1.831567e-22	203.0625
(x_1, x_2, x_4)	6.464166e-24	225
(x_1, x_3, x_4)	8.664767e-22	1491.8906

OVERFITTING

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Variables considered	J_{TRAIN}	J_{TEST}
(x_1, x_2)	3.750000e+01	100
(x_1, x_3)	2.403846e+01	959.8081
(x_1, x_4)	1.512500e+03	4900
(x_1, x_2, x_3)	1.831567e-22	203.0625
(x_1, x_2, x_4)	6.464166e-24	225
(x_1, x_3, x_4)	8.664767e-22	1491.8906

Best model

OVERFITTING

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Variables considered	J_{TRAIN}	J_{TEST}
(x_1, x_2)	3.750000e+01	100
(x_1, x_3)	2.403846e+01	959.8081
(x_1, x_4)	1.512500e+03	4900
(x_1, x_2, x_3)	1.831567e-22	203.0625
(x_1, x_2, x_4)	6.464166e-24	225
(x_1, x_3, x_4)	8.664767e-22	1491.8906

Best model

OVERFITTING

CV was also used here so set an hyper-parameter: «the number of variables to keep in the model».

Application on the example

So let us consider all models with either 2 or 3 variables (with at least the intercept each time).

By doing so, we add respectively 2 (ex: $\beta_2 = 0$ and $\beta_4 = 0$) or 1 constraint (idem).

For all these possible models, let us compute J_{TRAIN} and J_{TEST}:

Variables considered	J_{TRAIN}	J_{TEST}
(x_1, x_2)	3.750000e+01	100
(x_1, x_3)	2.403846e+01	959.8081
(x_1, x_4)	1.512500e+03	4900
(x_1, x_2, x_3)	1.831567e-22	203.0625
(x_1, x_2, x_4)	6.464166e-24	225
(x_1, x_3, x_4)	8.664767e-22	1491.8906

Best model

OVERFITTING

CV was also used here so set an hyper-parameter: «the number of variables to keep in the model».

Here apparently, keeping only 2 variables leads to the best model with the variable «Age», which was expected.

Overfitting can be handled with regularization.

Overfitting can be handled with regularization.

Cross-Validation can both help to:

Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**

Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**

Overfitting, Cross-Validation & Regularization

Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**

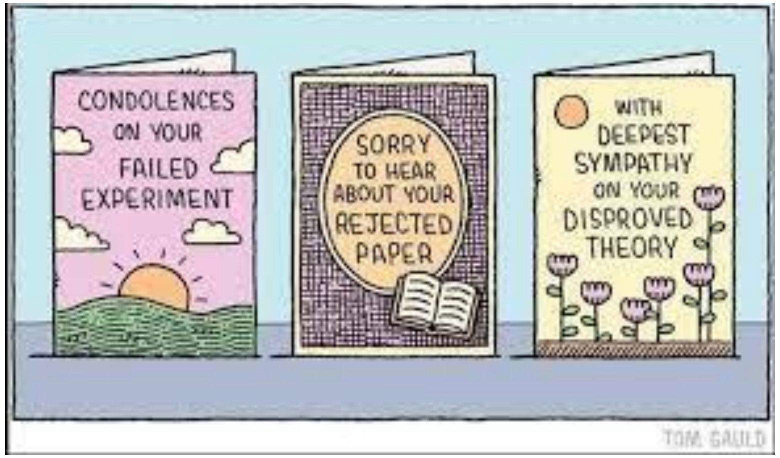
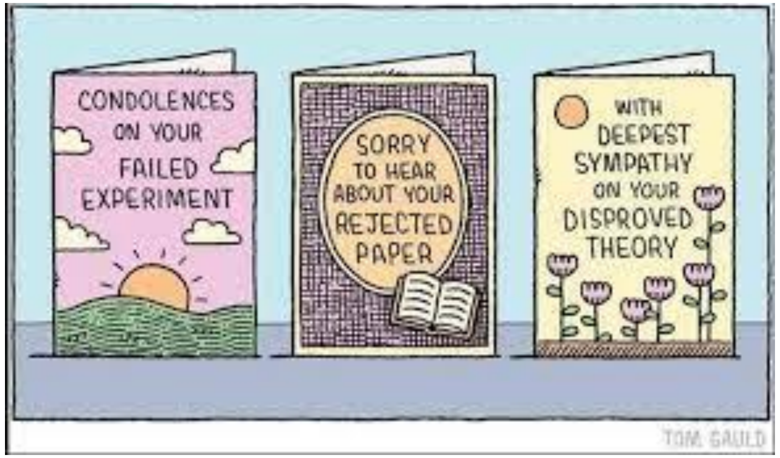


Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

Overfitting, Cross-Validation & Regularization

Overfitting can be handled with regularization.

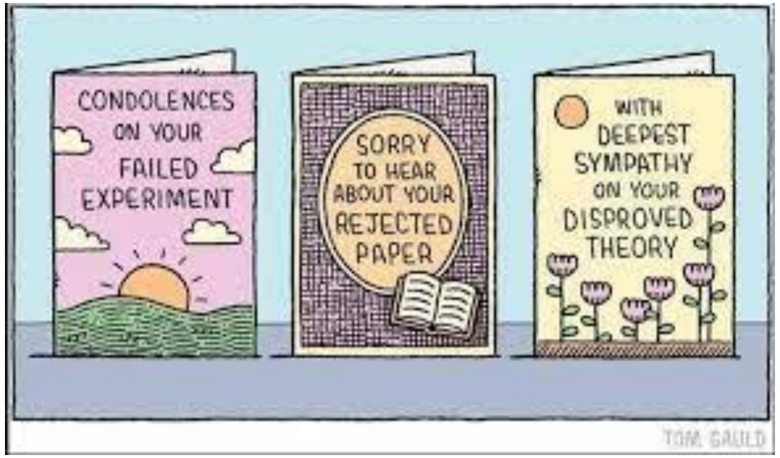
Cross-Validation can both help to:
1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**



Classical mistake to avoid with Cross-Validation: «**Double Dipping**».

Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

Overfitting, Cross-Validation & Regularization



Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**

Classical mistake to avoid with Cross-Validation: «**Double Dipping**».

➔ The whole point of Cross-Validation is to keep the train and the test sets **independant** from each other.

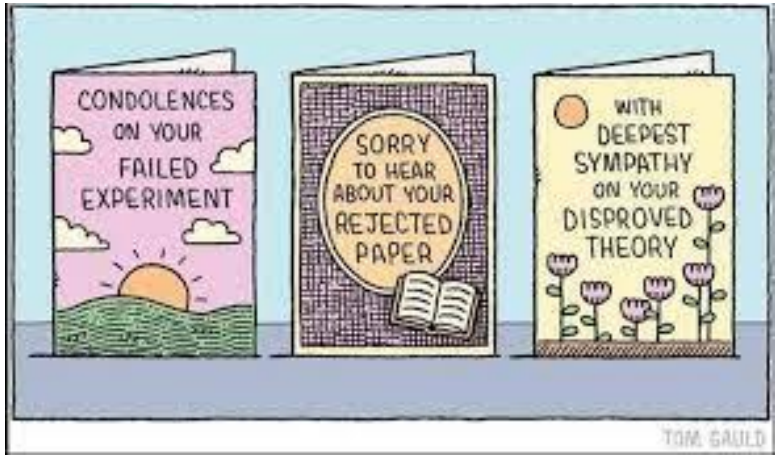
Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

Overfitting, Cross-Validation & Regularization

Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**



Classical mistake to avoid with Cross-Validation: «**Double Dipping**».

➔ The whole point of Cross-Validation is to keep the train and the test sets **independant** from each other.

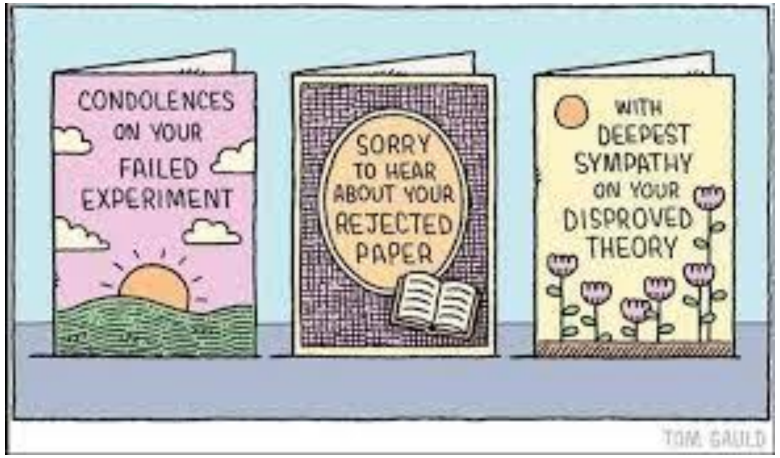
This is no longer the case when for example:

Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**



Classical mistake to avoid with Cross-Validation: «**Double Dipping**».

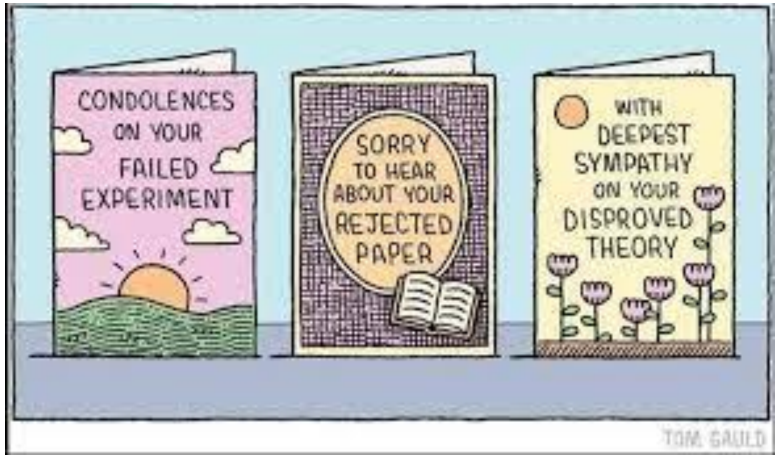
➔ The whole point of Cross-Validation is to keep the train and the test sets **independant** from each other.

This is no longer the case when for example:

1. Normalization accross subjects is performed on the whole data-set.

Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

Overfitting, Cross-Validation & Regularization



Overfitting can be handled with regularization.

Cross-Validation can both help to:

1. **realize if the model overfits or not**
2. **tune the hyper-parameters (associated with the regularization).**

Classical mistake to avoid with Cross-Validation: «**Double Dipping**».

➔ The whole point of Cross-Validation is to keep the train and the test sets **independant** from each other.

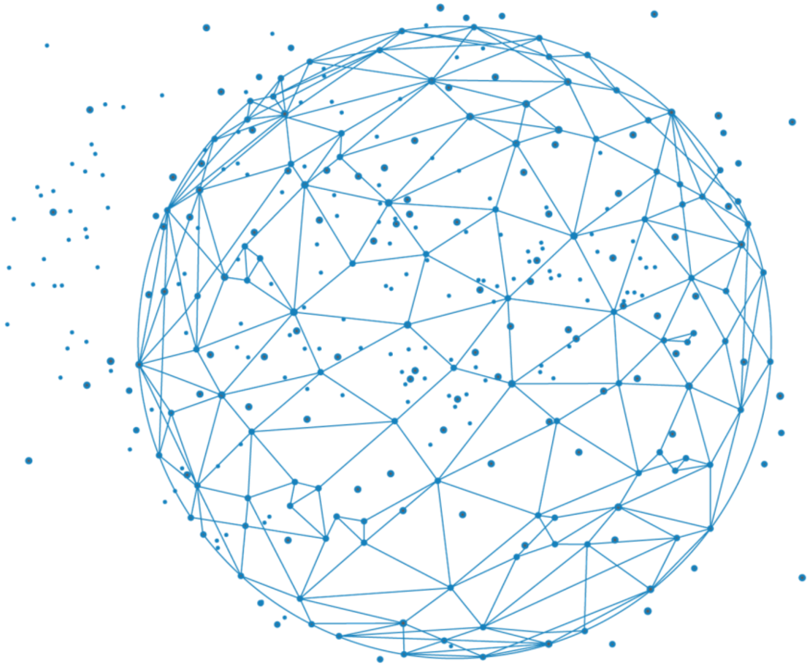
This is no longer the case when for example:

1. Normalization accross subjects is performed on the whole data-set.
2. Variable selection is performed on the whole data-set (ex: differentially expressed genes)

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Figure taken from https://typeset.io/resources/top-reasons-for-research-paper-rejection/

How do we regularize CCA ?



Two-blocks special cases: PLS & CCA

Canonical Correlation Analysis (CCA)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Partial Least Squares (PLS2)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Canonical Correlation Analysis (CCA)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i\mathbf{w}_i)=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Partial Least Squares (PLS2)

$$\max_{\substack{\mathbf{w}_1,\mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2=1}} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Regularized-CCA

$$\max_{\mathbf{w}_1,\mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } (1-\tau_i)\mathrm{Var}(\mathbf{X}_i\mathbf{w}_i) + \tau_i\|\mathbf{w}_i\|_2^2 = 1.$$

Canonical Correlation Analysis (CCA)

$$\max_{\substack{\mathbf{w}_1, \mathbf{w}_2 \\ \mathrm{Var}(\mathbf{X}_i \mathbf{w}_i)=1}} \mathrm{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

Partial Least Squares (PLS2)

$$\max_{\substack{\mathbf{w}_1, \mathbf{w}_2 \\ \|\mathbf{w}_i\|_2^2=1}} \mathrm{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

Regularized-CCA

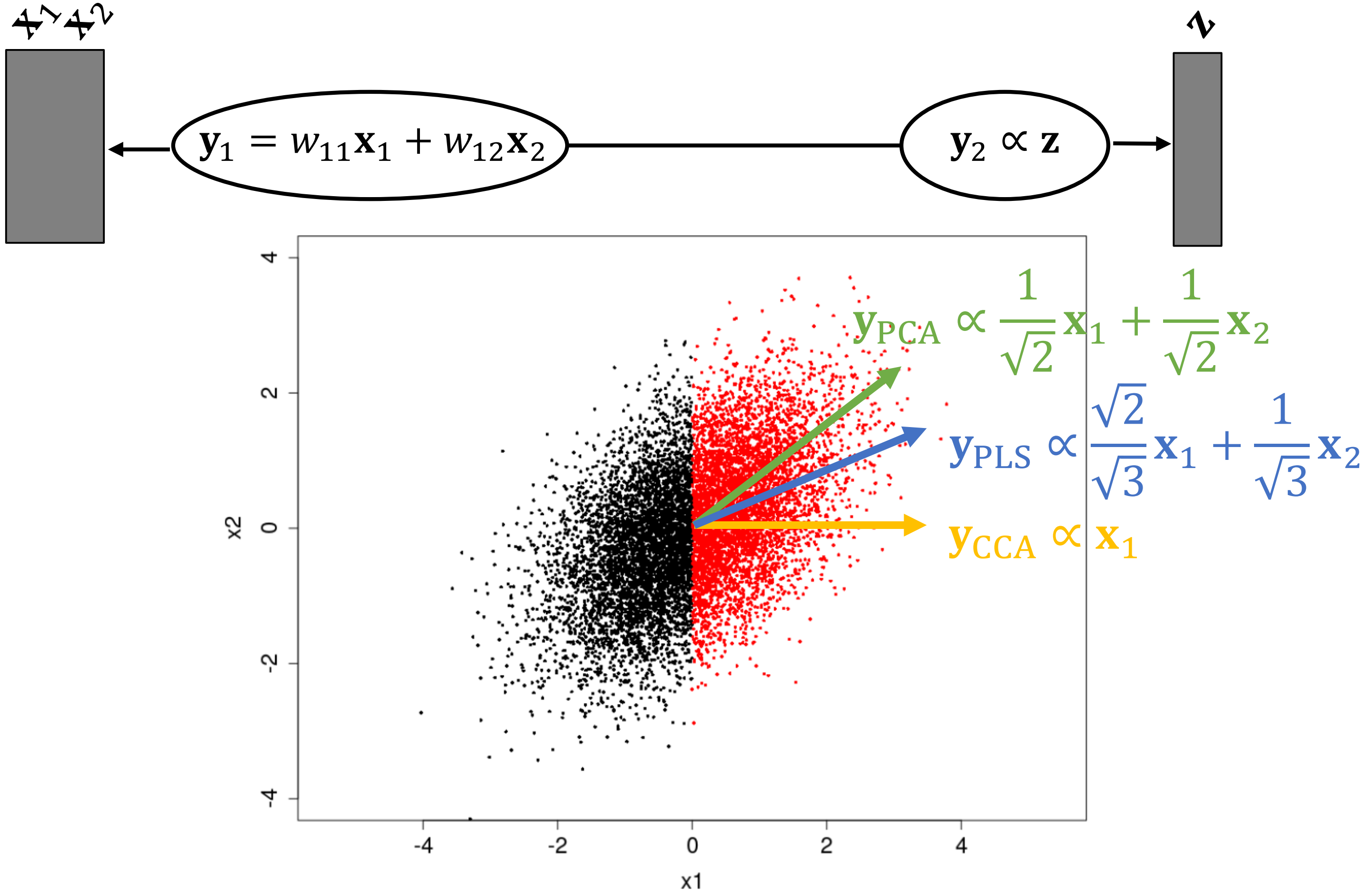
$$\max_{\mathbf{w}_1, \mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

$$\mathrm{s.t.}\ (1 - \tau_i)\mathrm{Var}(\mathbf{X}_i \mathbf{w}_i) + \tau_i \|\mathbf{w}_i\|_2^2 = 1.$$

PLS & CCA with a figure

$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

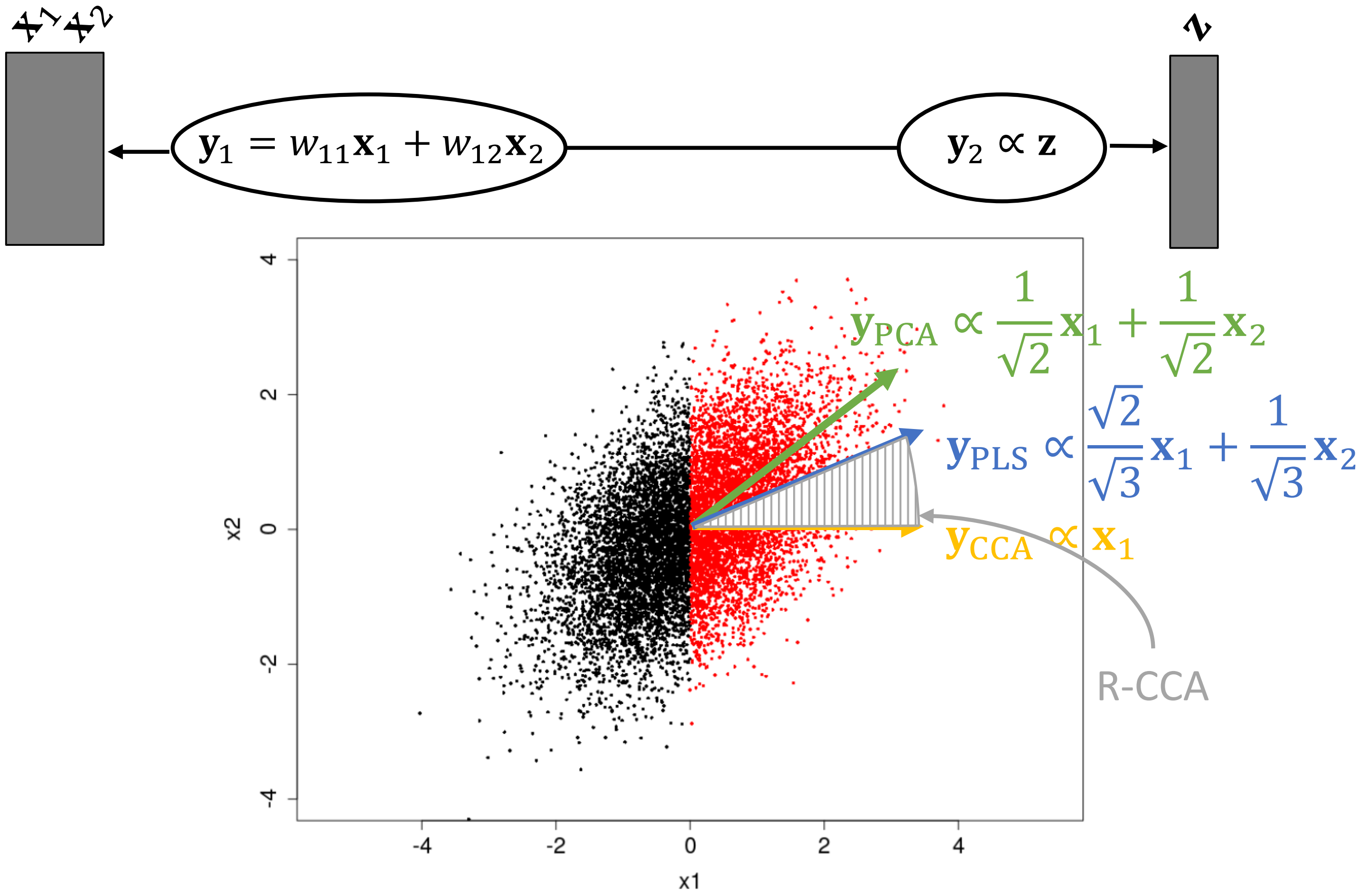
$$(\mathbf{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



PLS & CCA with a figure

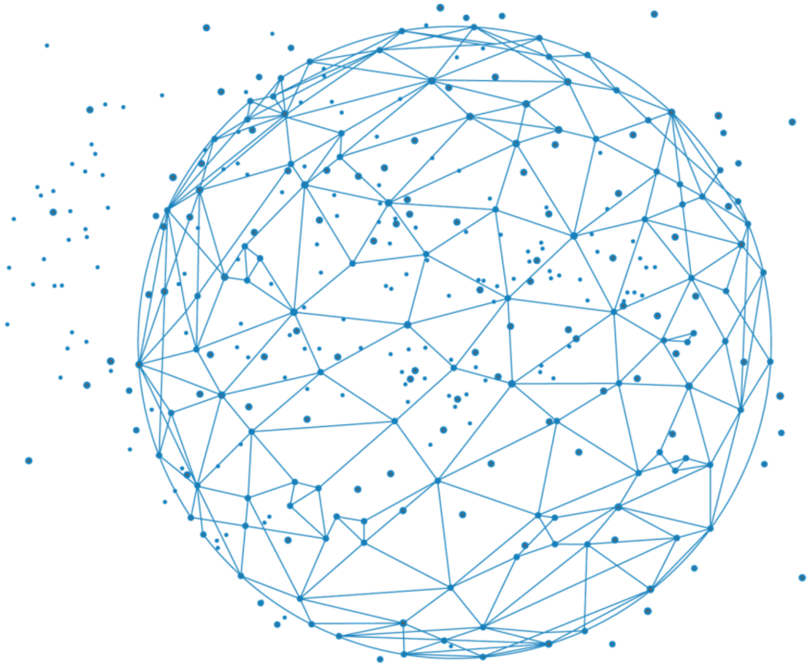
$$[\mathbf{x}_1\ \mathbf{x}_2] \sim \mathcal{N}\left((0,0), \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$

$$(\mathbf{z})_i = \begin{cases} 0 \text{ if } (\mathbf{x})_i < 0 \\ 1 \text{ otherwise} \end{cases}$$



Let us see how Regularize CCA performs on the MDD case study

➔ See section 2.4 & 2.5 on the Rmarkdown `MDD_case_study_RGCCA`



1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

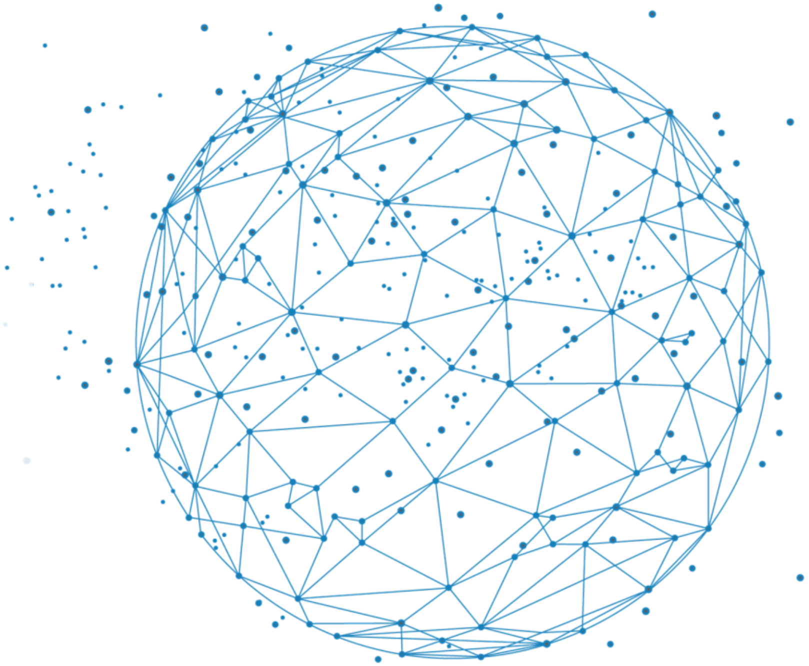
4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)

5. Supervised analysis with RGCCA

6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)

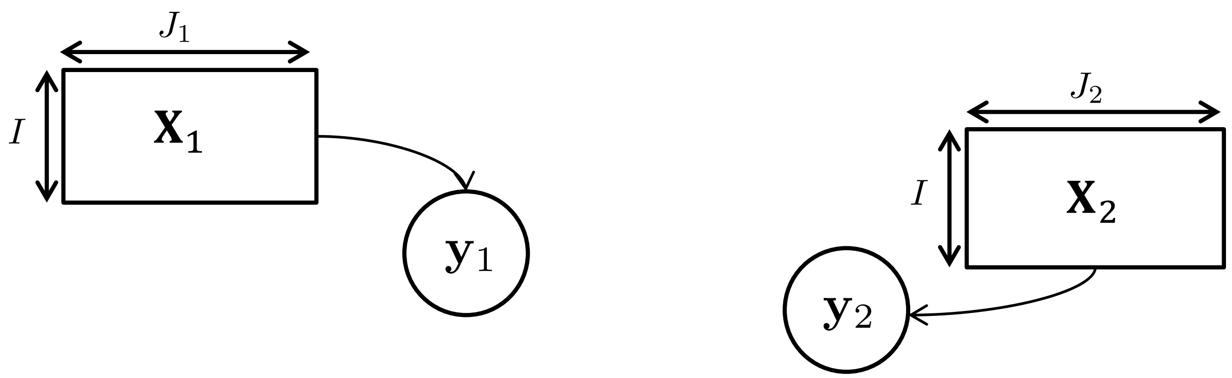
7. The flexible Optimization Framework of RGCCA

   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

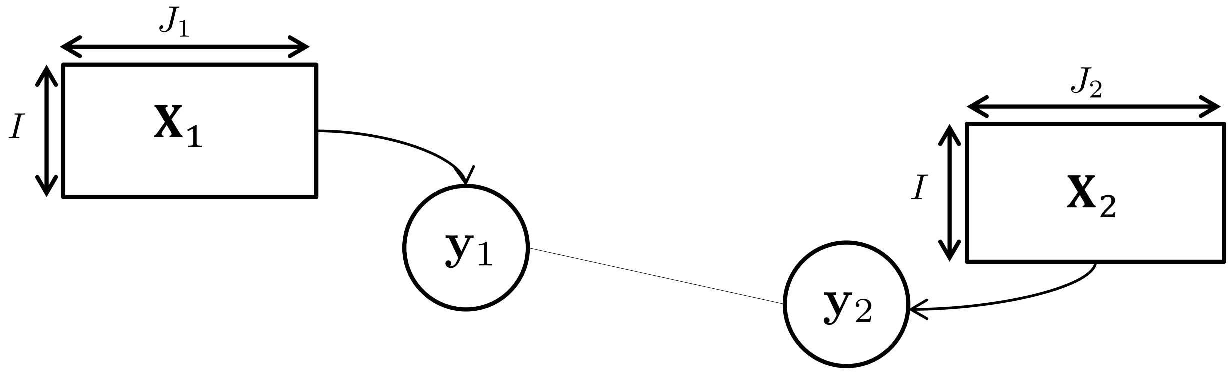


Regularized Generalized Canonical Correlation Analysis (RGCCA)

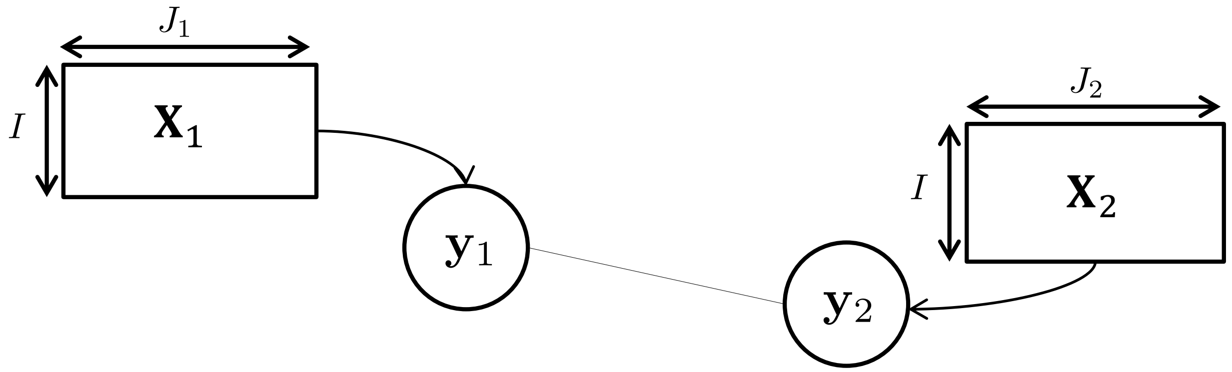
Regularized Generalized Canonical Correlation Analysis (RGCCA)



Regularized Generalized Canonical Correlation Analysis (RGCCA)

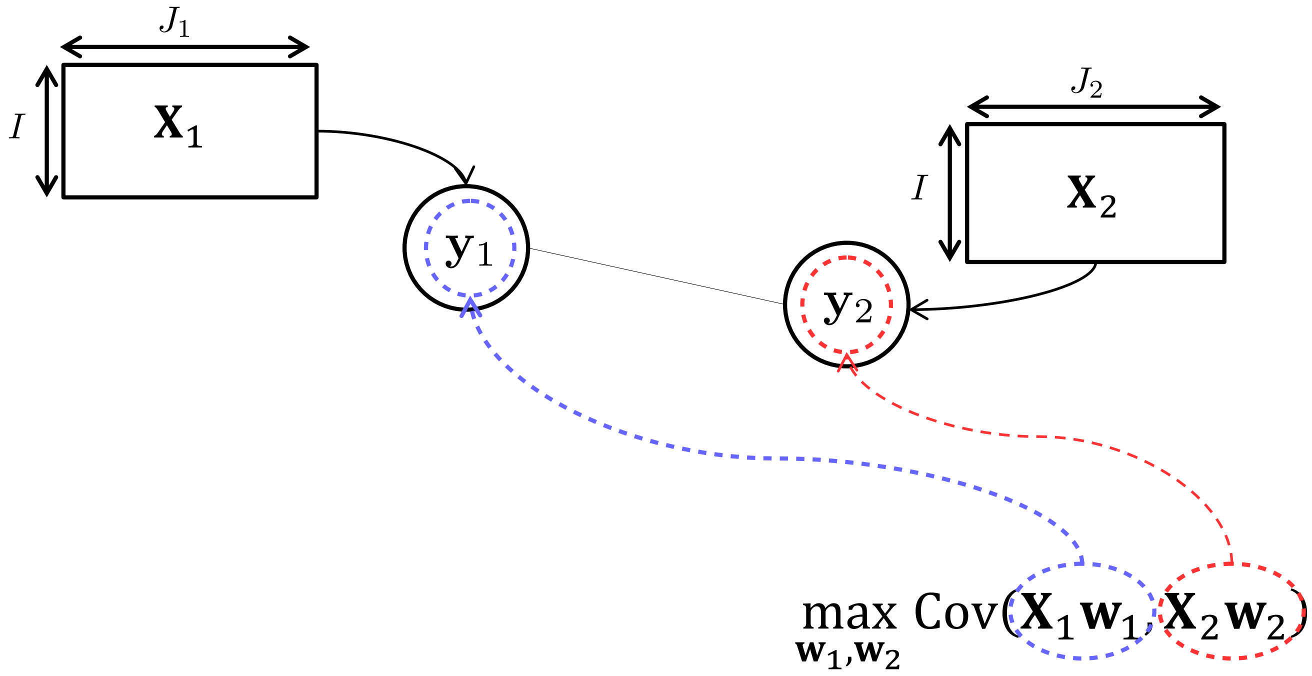


Regularized Generalized Canonical Correlation Analysis (RGCCA)



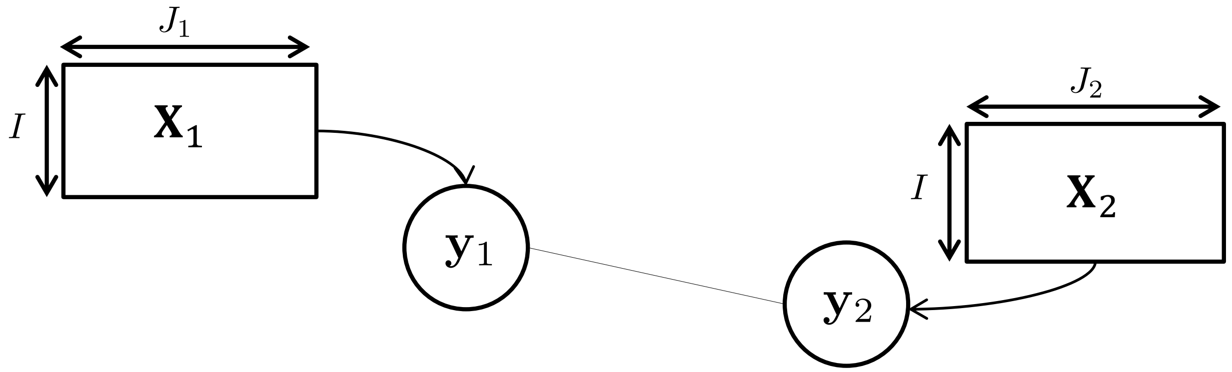
$$\max_{\mathbf{w}_1,\mathbf{w}_2} \text{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

Regularized Generalized Canonical Correlation Analysis (RGCCA)



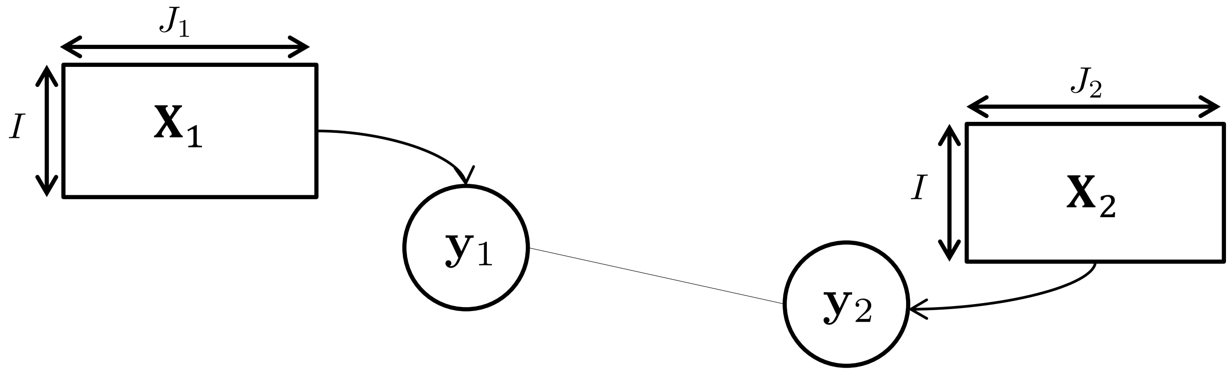
INSTITUT FRANÇAIS DE BIOINFORMATIQUE IFB elixir FRANCE

Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1, \mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

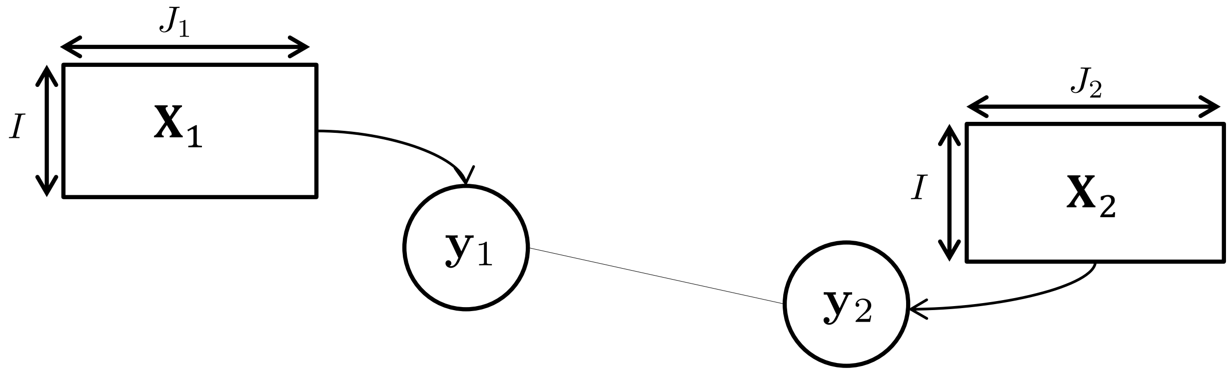
Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1, \mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t.}\ \mathbf{w}_l^\top \mathbf{X}_l^\top \mathbf{X}_l \mathbf{w}_l = I, \qquad l = 1, 2.$$

Regularized Generalized Canonical Correlation Analysis (RGCCA)

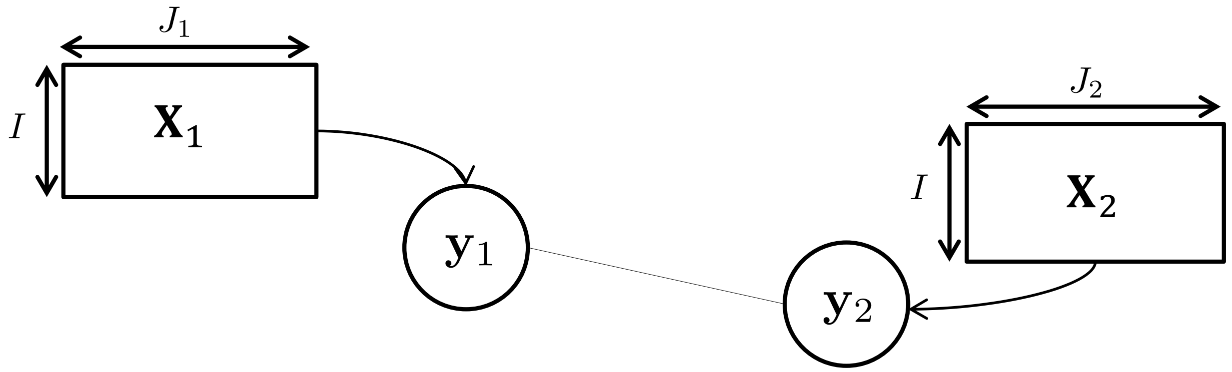


$$\max_{\mathbf{w}_1,\mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t.}\ \mathbf{w}_l^\top \mathbf{X}_l^\top \mathbf{X}_l \mathbf{w}_l = I, \qquad l = 1, 2.$$

➡ Canonical Correlation Analysis

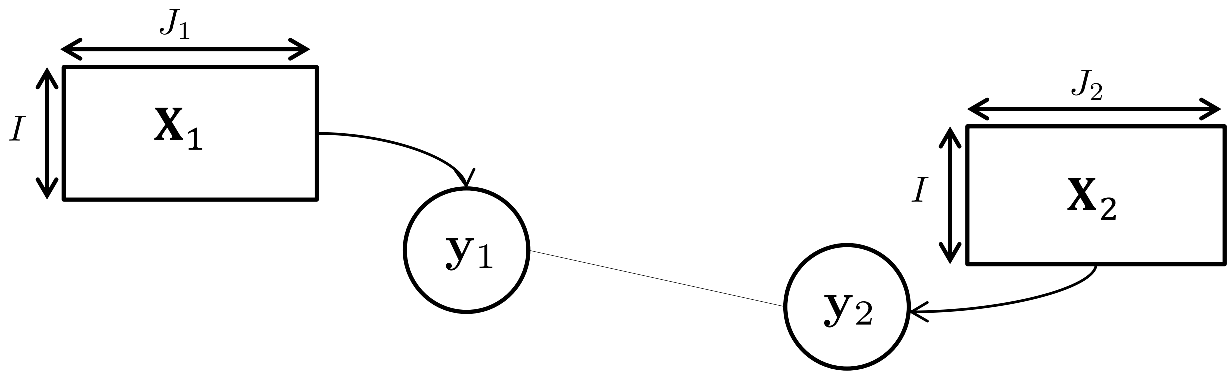
Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t.}\ \mathbf{w}_l^\top \mathbf{X}_l^\top \mathbf{X}_l \mathbf{w}_l = I, \qquad l = 1, 2.$$

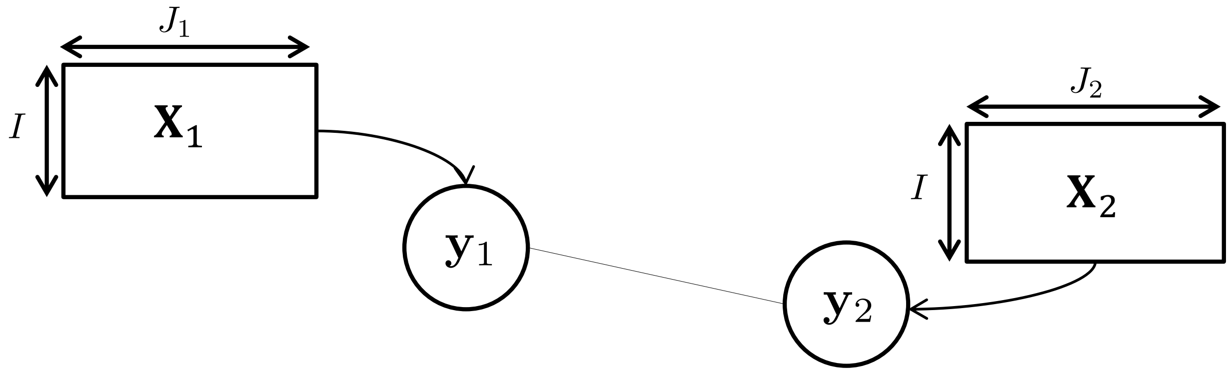
Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1, \mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t.}\ \|\mathbf{w}_l\|_2^2 = 1, \qquad l = 1, 2.$$

Regularized Generalized Canonical Correlation Analysis (RGCCA)

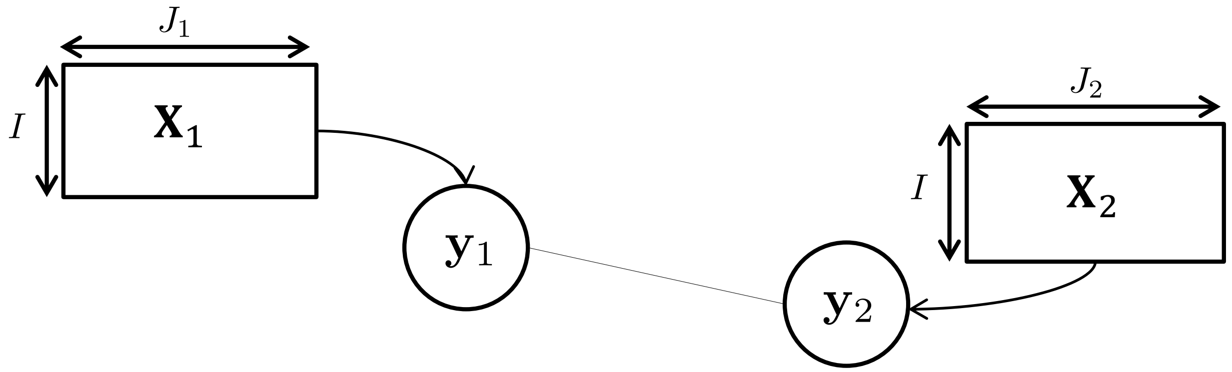


$$\max_{\mathbf{w}_1,\mathbf{w}_2} \operatorname{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } \|\mathbf{w}_l\|_2^2 = 1, \qquad l = 1, 2.$$

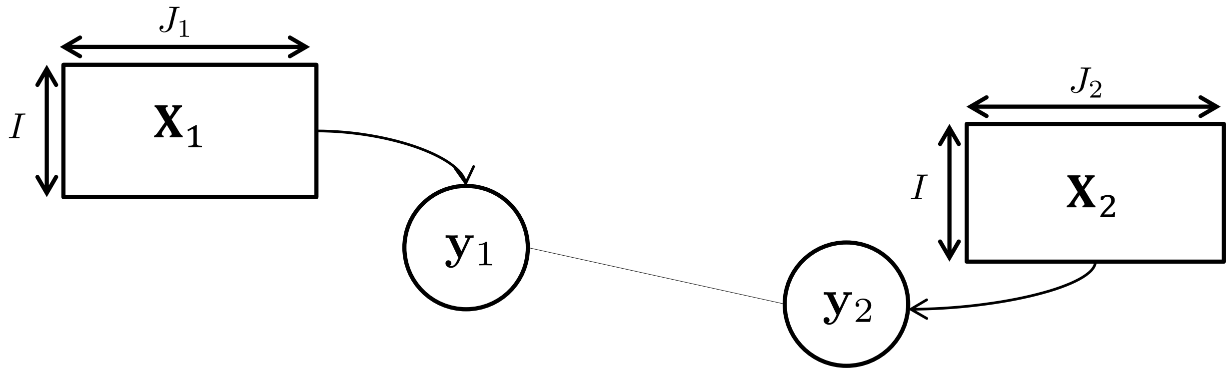
➡ Partial Least Squares 2

Regularized Generalized Canonical Correlation Analysis (RGCCA)



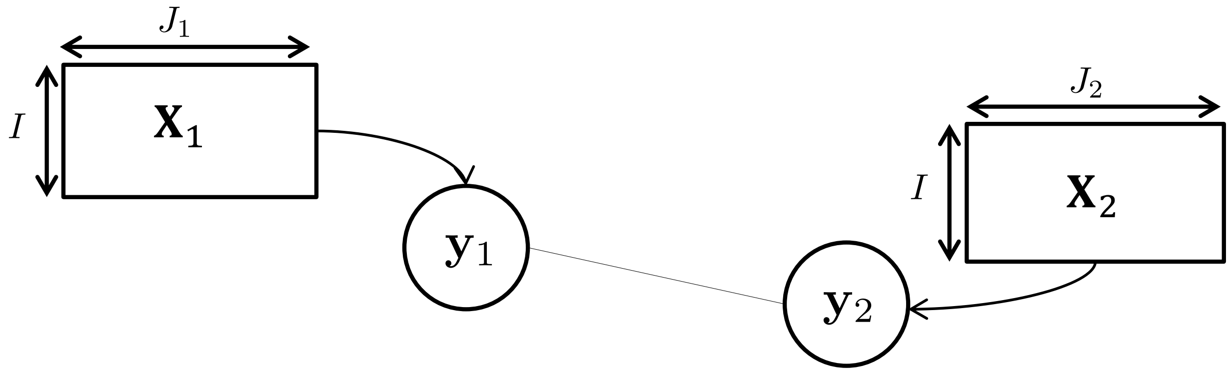
$$\max_{\mathbf{w}_1,\mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } \|\mathbf{w}_l\|_2^2 = 1, \qquad l = 1, 2.$$



$$\max_{\mathbf{w}_1,\mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \qquad l = 1, 2.$$

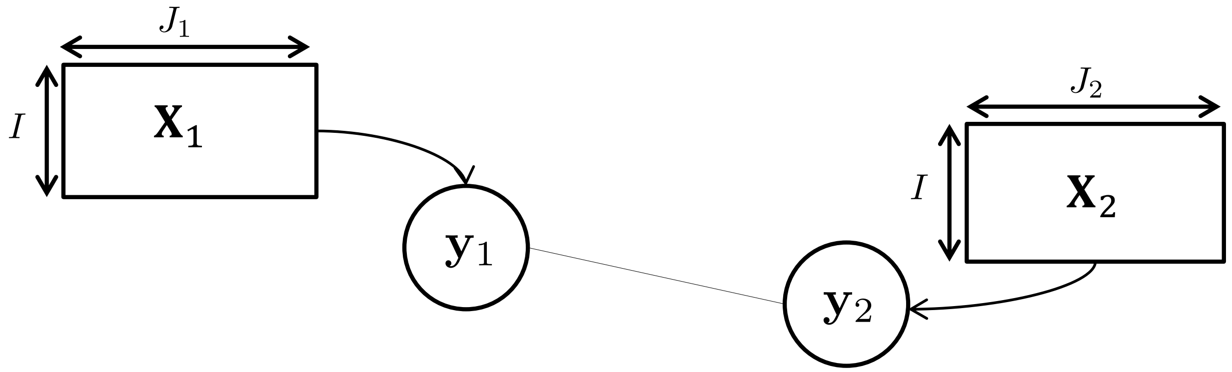


$$\max_{\mathbf{w}_1,\mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \qquad l = 1, 2.$$

where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix

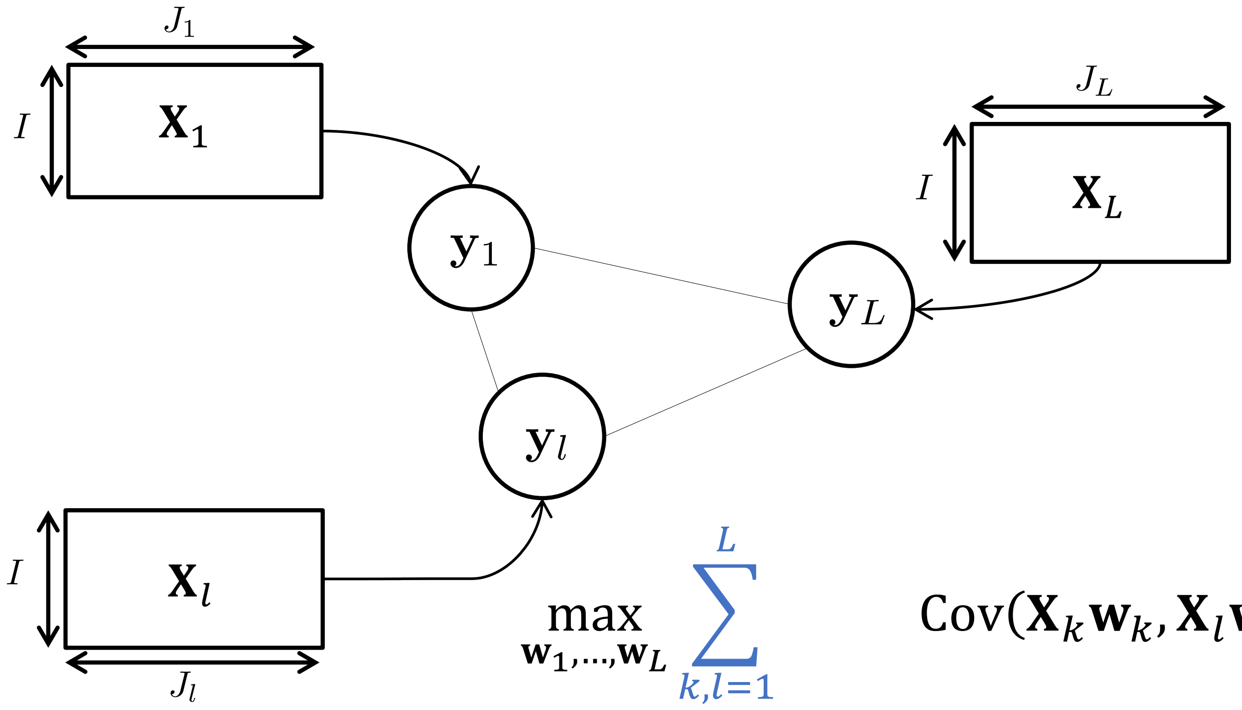
Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1, \mathbf{w}_2} \mathrm{Cov}(\mathbf{X}_1\mathbf{w}_1, \mathbf{X}_2\mathbf{w}_2)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \qquad l = 1, 2.$$

Regularized Generalized Canonical Correlation Analysis (RGCCA)



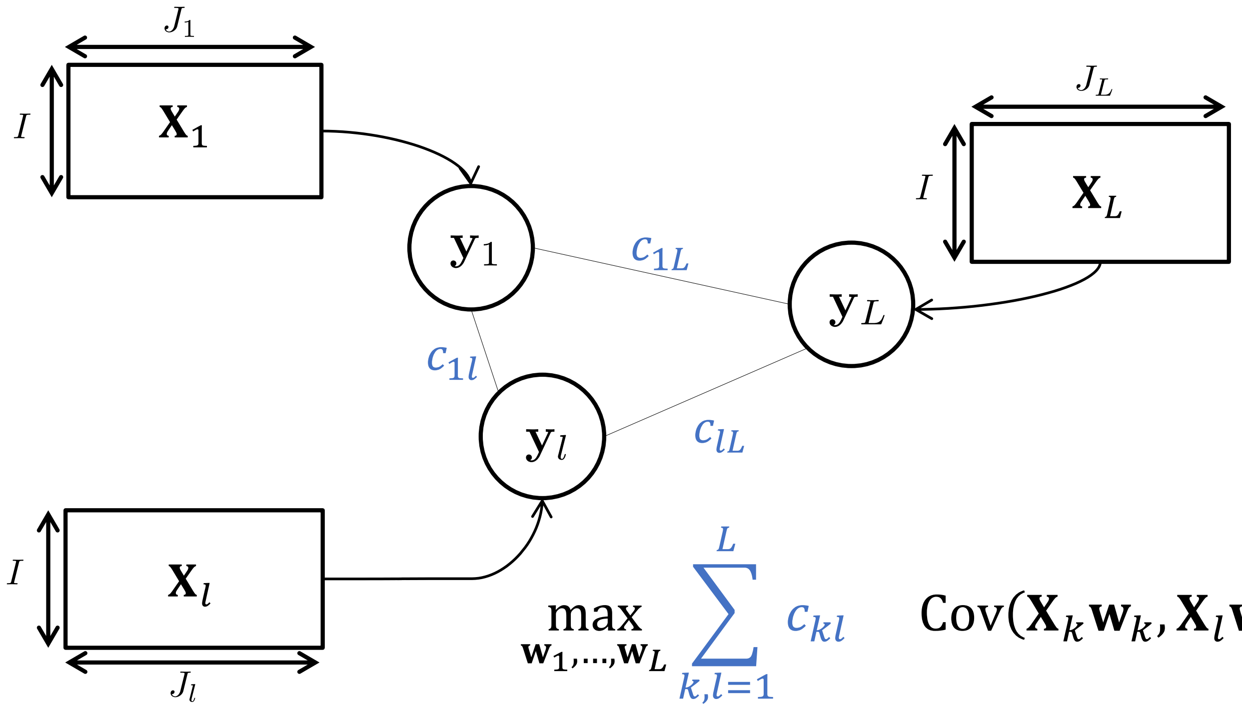
$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} \operatorname{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ l = 1, \ldots, L.$$

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

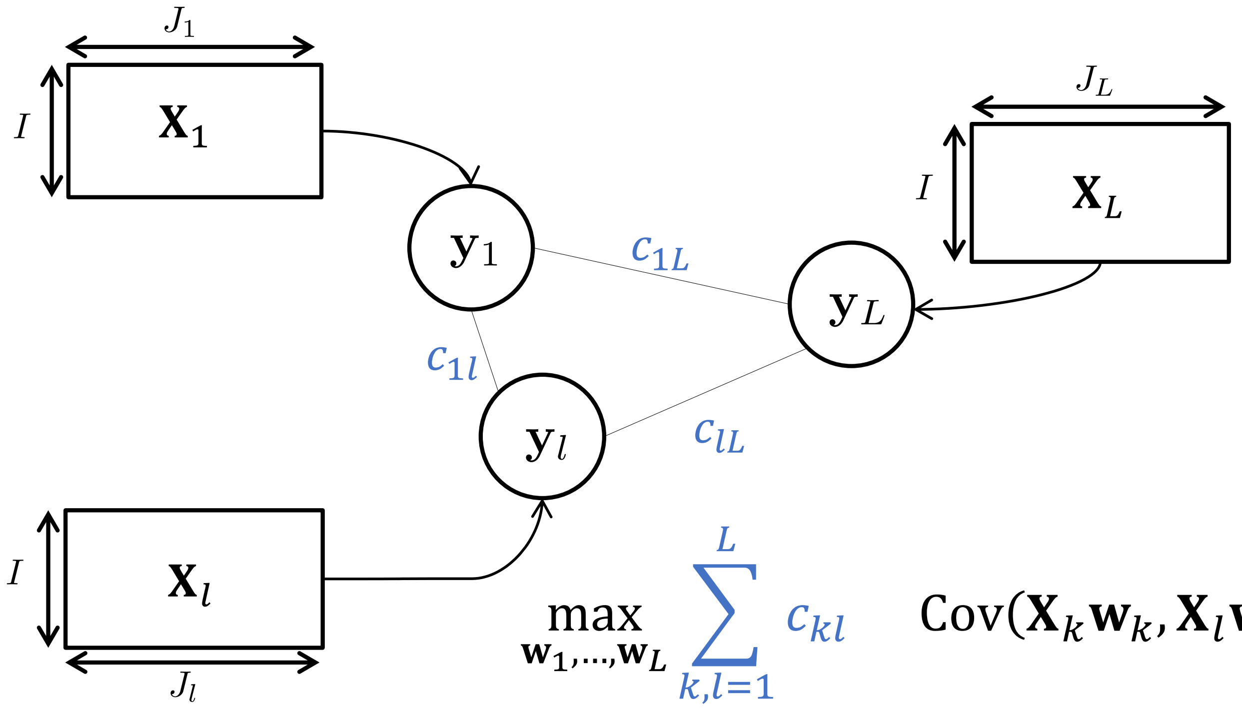
eli𝑥ir FRANCE

Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl} \quad \mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)$$

$$\text{s.t.}\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

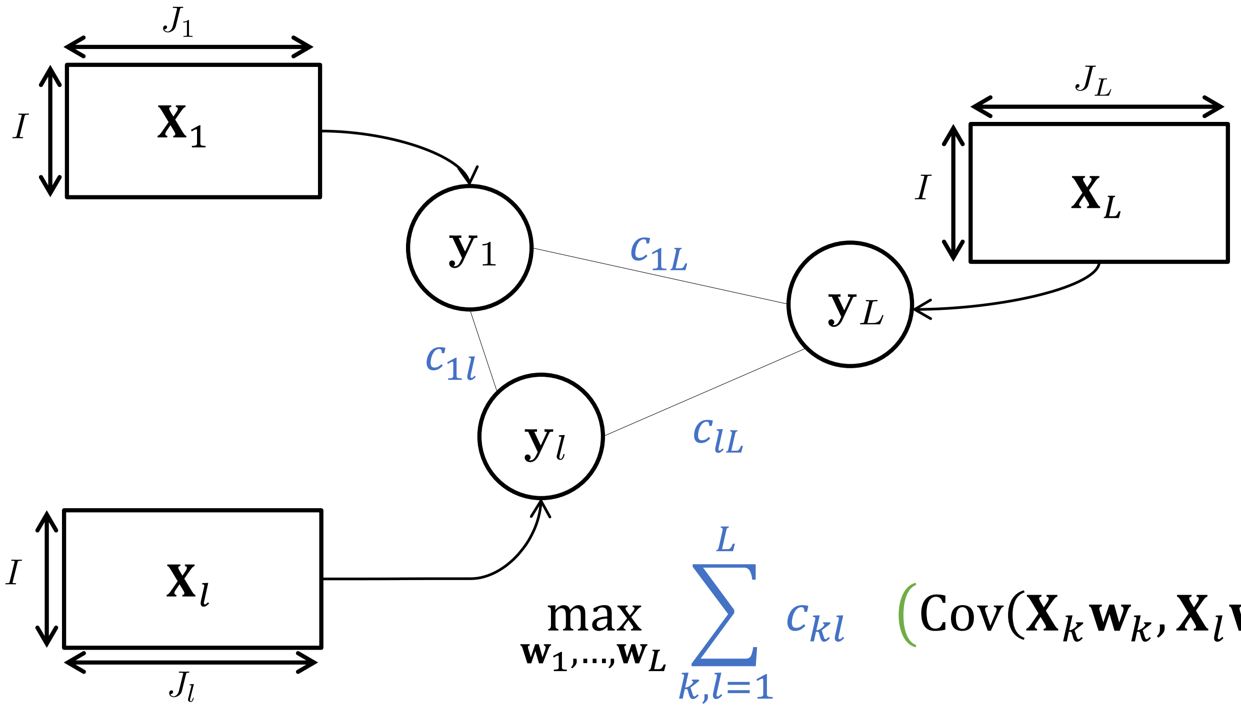


$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl} \quad \mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

⇨ if all blocks are connected and $\mathbf{M}_l = \mathbf{I}_l$ ⇨ SUMCOV-2

Regularized Generalized Canonical Correlation Analysis (RGCCA)

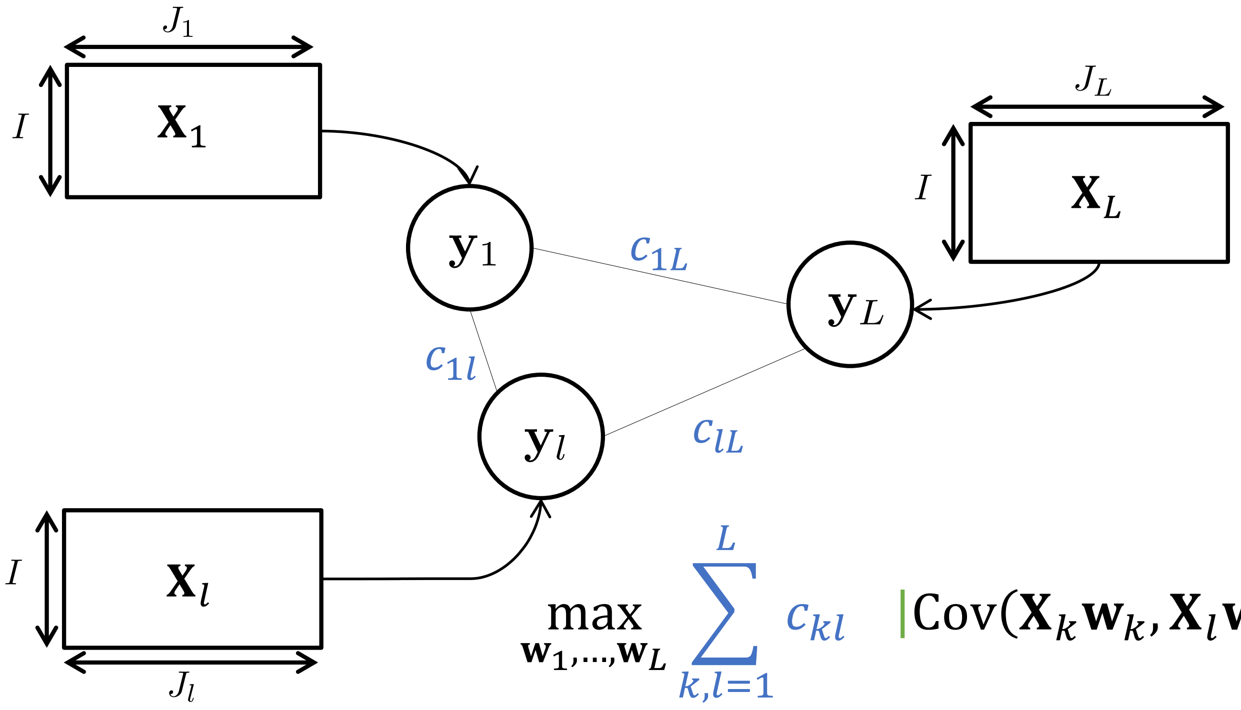


$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl} \left(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\right)^2$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \; l = 1, \ldots, L.$$

➡ if all blocks are connected and $\mathbf{M}_l = \mathbf{I}_l$ ➡ SSQCOV-2

Regularized Generalized Canonical Correlation Analysis (RGCCA)

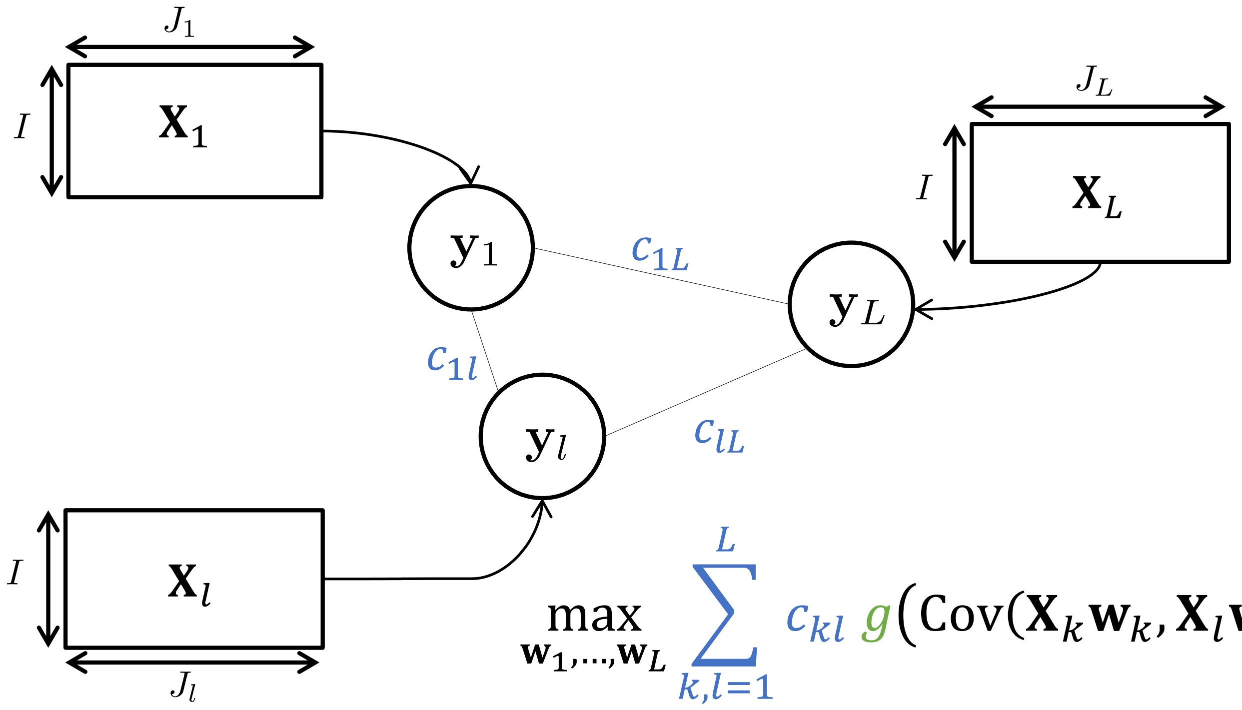


$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl} \left|\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\right|$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \; l = 1, \ldots, L.$$

if all blocks are connected and $\mathbf{M}_l = \mathbf{I}_l$ SABSCOV-2

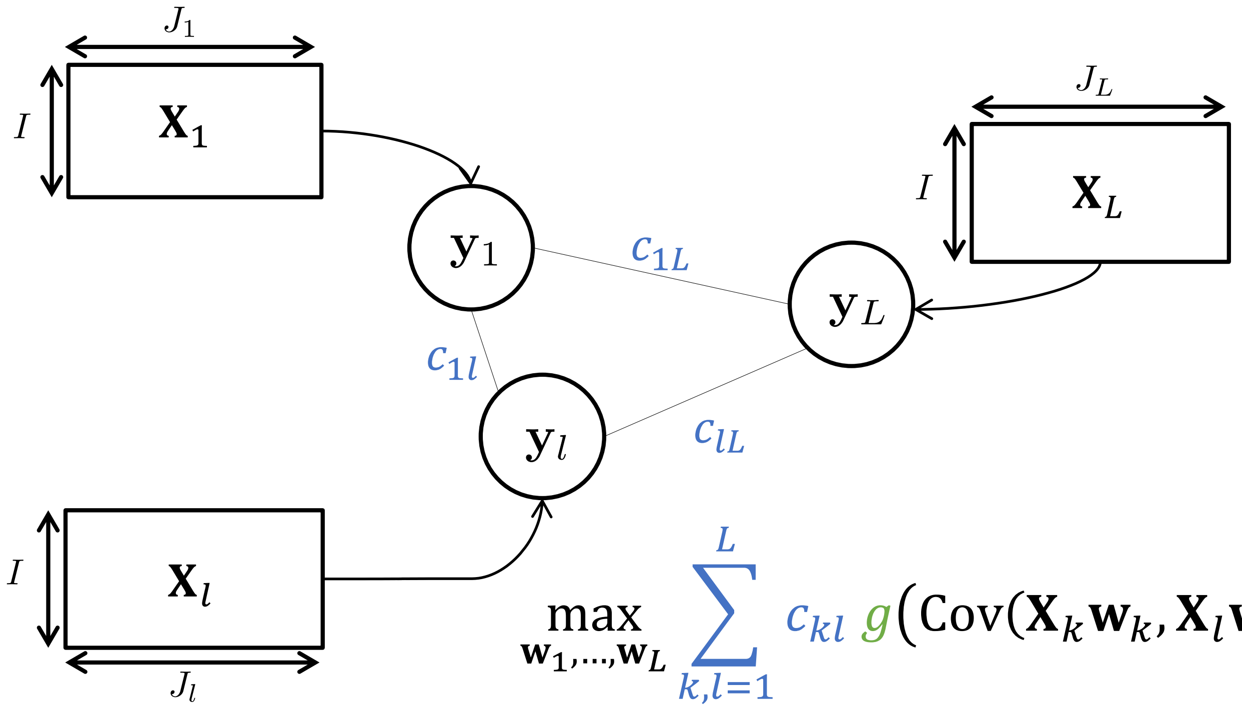
Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ l = 1, \ldots, L.$$

Regularized Generalized Canonical Correlation Analysis (RGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ l = 1, \ldots, L.$$

➡ with g a continuous, convex and derivable function.

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\; g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE elixir FRANCE

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\mathrm{s.t.}\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

➡ With "g" a continuous, convex and derivable function.

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.}\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ l = 1, \ldots, L.$$

- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\left(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\right)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

elixir FRANCE

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix.

Most of the time (this is the case today !) $\mathbf{M}_l$ is chosen such that:

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \; l = 1, \ldots, L.$$

- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix.

Most of the time (this is the case today !) $\mathbf{M}_l$ is chosen such that:

$$\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = \quad (1-\tau_l)\mathrm{Var}(\mathbf{X}_l\mathbf{w}_l) + \tau_l \|\mathbf{w}_l\|_2^2 \quad = 1.$$

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1, \; l = 1, \dots, L.$$

- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix.

Most of the time (this is the case today !) $\mathbf{M}_l$ is chosen such that:

$$\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = \mathbf{w}_l^\top \left((1-\tau_l)I^{-1}\mathbf{X}_l^\top\mathbf{X}_l + \tau_l \mathbf{I}_{J_l}\right)\mathbf{w}_l = 1.$$

Summary of RGCCA

The Regularized Generalized Canonical Correlation Analysis (RGCCA) Optimization criterion :

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t. } \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1,\ \ l = 1, \ldots, L.$$

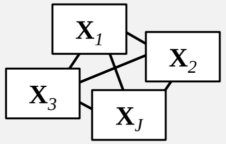
- With "g" a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where $\mathbf{M}_l$ is any $J_l \times J_l$ positive definite matrix.

Most of the time (this is the case today !) $\mathbf{M}_l$ is chosen such that:

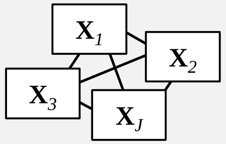
$$\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = \mathbf{w}_l^\top \underbrace{\left((1-\tau_l)I^{-1}\mathbf{X}_l^\top \mathbf{X}_l + \tau_l \mathbf{I}_{J_l}\right)}_{\text{Regularized version of the sample covariance matrix}} \mathbf{w}_l = 1.$$

Regularized version of the sample covariance matrix

Overview of the Multi-Block litterature

ALL BLOCKS ARE INTERCONNECTED 	
SUMCOR (Horst, 1961)	$\max_{\mathbf{w}_j} \sum_{j,k} \text{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOR (Kettenring, 1961)	$\max_{\mathbf{w}_j} \sum_{j,k} \text{cor}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOR (Wold, 1982)	$\max_{\mathbf{w}_j} \sum_{j,k} \lvert\text{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$

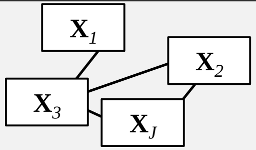
Overview of the Multi-Block litterature

ALL BLOCKS ARE INTERCONNECTED	
SUMCOR (Horst, 1961)	$\max_{\mathbf{w}_j} \sum_{j,k} \mathrm{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOR (Kettenring, 1961)	$\max_{\mathbf{w}_j} \sum_{j,k} \mathrm{cor}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOR (Wold, 1982)	$\max_{\mathbf{w}_j} \sum_{j,k} \lvert\mathrm{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$
SUMCOV (Van de Geer, 1984)	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} \mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOV (Hanafi & Kiers, 2006)	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} \mathrm{cov}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOV (Krämer, 2007)	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} \lvert\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$

Overview of the Multi-Block litterature

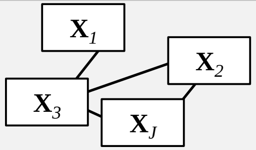
BLOCKS ARE PARTIALLY CONNECTED

$c_{jk} = 1$ if $\mathbf{X}_j \leftrightarrow \mathbf{X}_k$, 0 otherwise



SUMCOR	$\max_{\mathbf{w}_j} \sum_{j,k} c_{jk} \text{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOR	$\max_{\mathbf{w}_j} \sum_{j,k} c_{jk} \text{cor}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOR	$\max_{\mathbf{w}_j} \sum_{j,k} c_{jk} \lvert\text{cor}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$
SUMCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \lvert\text{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$

Overview of the Multi-Block litterature

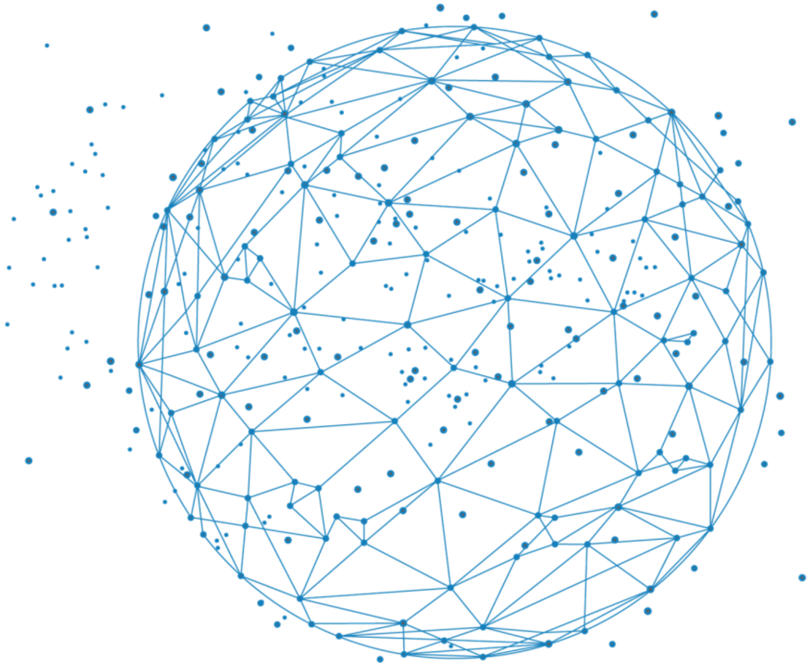


BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1$ if $\mathbf{X}_j \leftrightarrow \mathbf{X}_k$, 0 otherwise	
SUMCOR	$\max_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum_{j,k} c_{jk} \mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOR	$\max_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum_{j,k} c_{jk} \mathrm{cov}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOR	$\max_{\mathrm{var}(\mathbf{X}_j\mathbf{w}_j)=1} \sum_{j,k} c_{jk} \lvert\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$
SUMCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SSQCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \mathrm{cov}^2(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)$
SABSCOV	$\max_{\lVert\mathbf{w}_j\rVert=1} \sum_{j,k} c_{jk} \lvert\mathrm{cov}(\mathbf{X}_j\mathbf{w}_j, \mathbf{X}_k\mathbf{w}_k)\rvert$

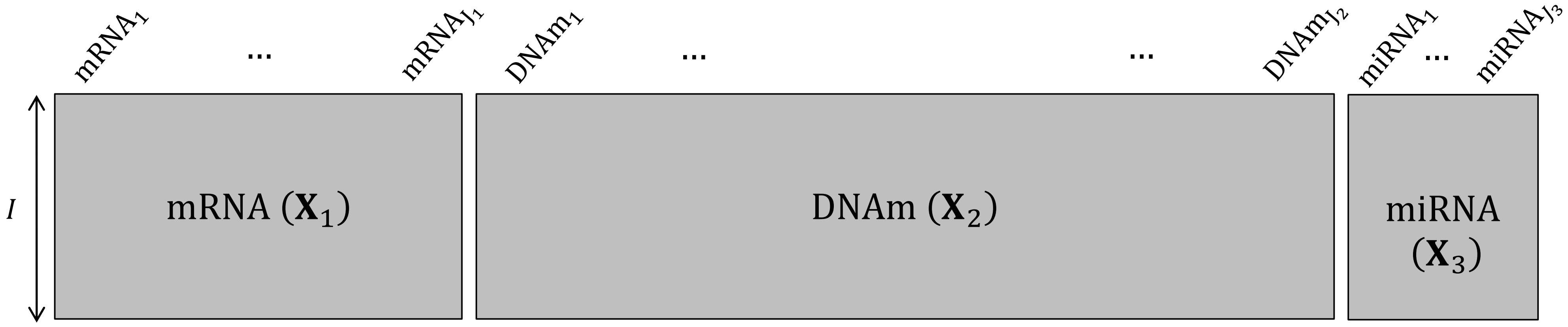
Courtesy to Arthur Tenenhaus.

Let us see how RGCCA performs on the MDD case study

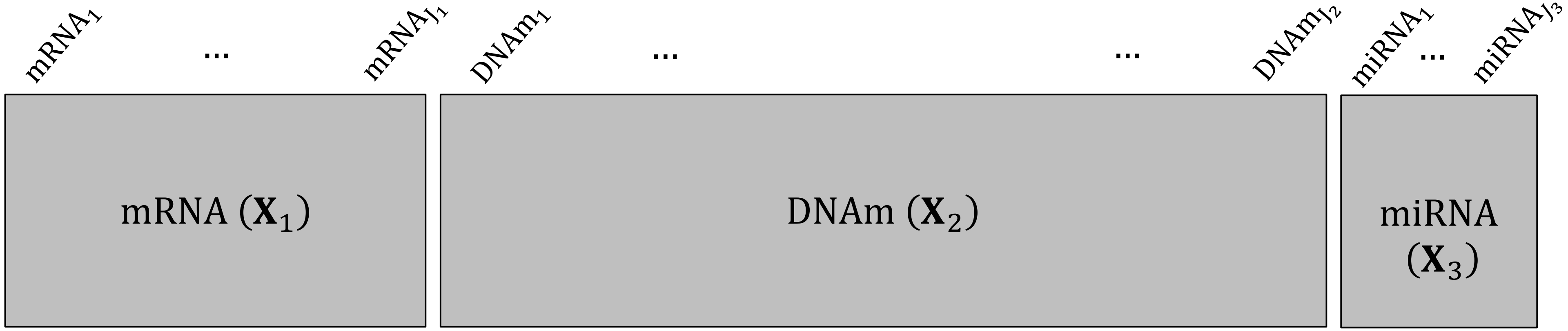
➔ See section 3.2 on the Rmarkdown `MDD_case_study_RGCCA`



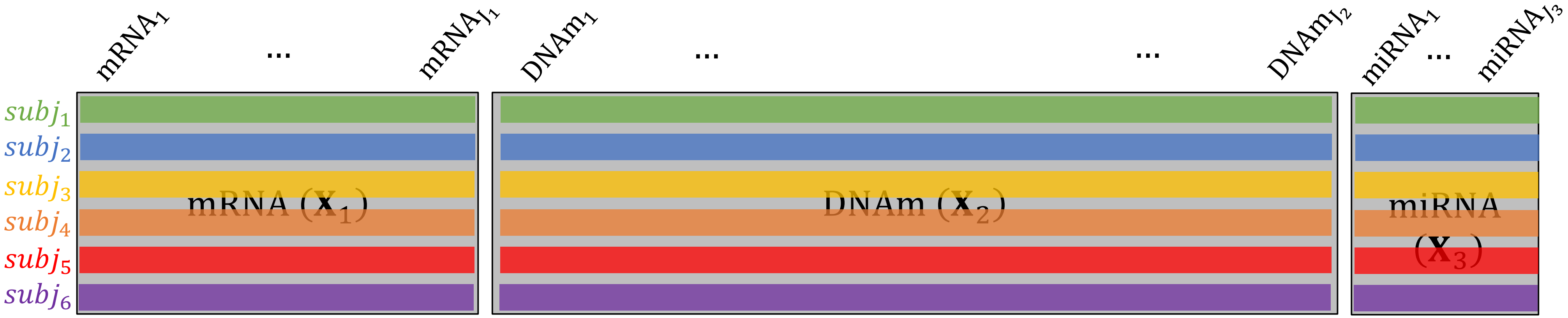
Tune parameters in an unsupervised setting



Tune parameters in an unsupervised setting



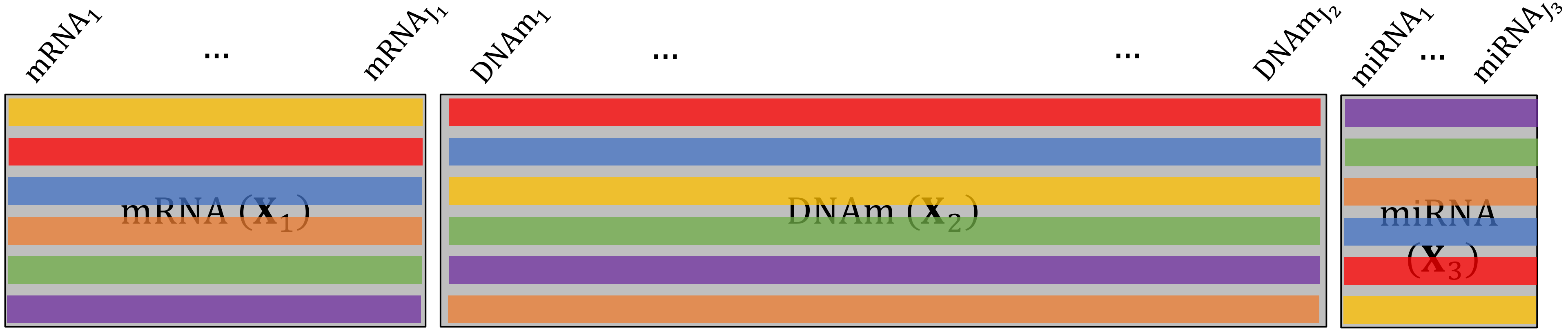
Tune parameters in an unsupervised setting



Tune parameters in an unsupervised setting

Permutation n°1

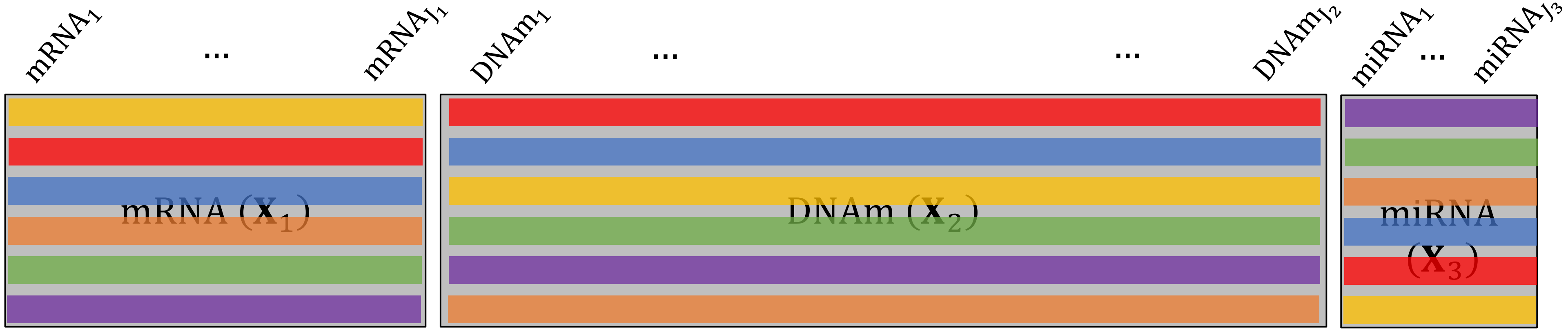
Tune parameters in an unsupervised setting



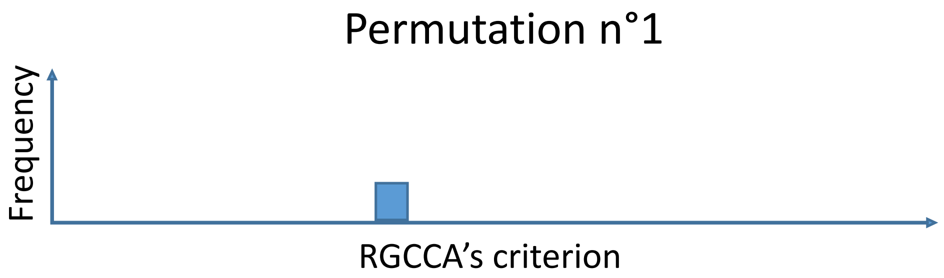
Permutation n°1

Parameter set n°1

Tune parameters in an unsupervised setting



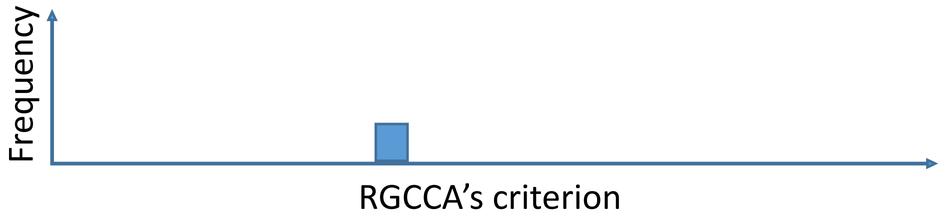
Parameter set n°1



Tune parameters in an unsupervised setting

Permutation n°1

Parameter set n°1



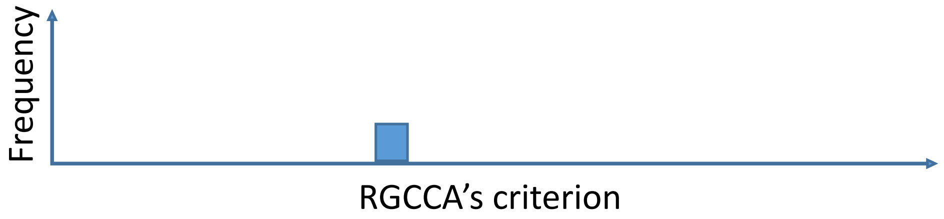
⋮

Parameter set n°K

Tune parameters in an unsupervised setting

Permutation n°1

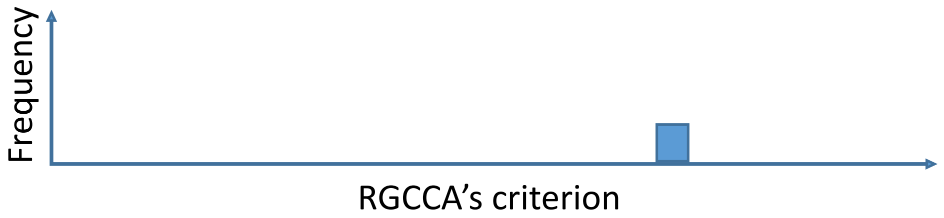
Parameter set n°1



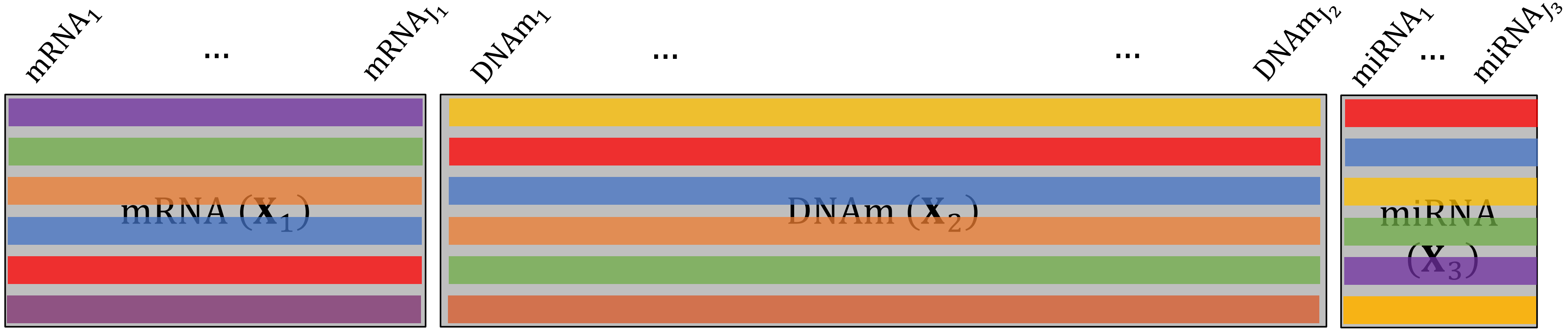
⋮

⋮

Parameter set n°K

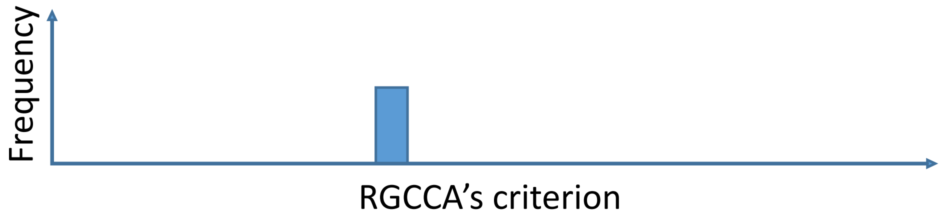


Tune parameters in an unsupervised setting



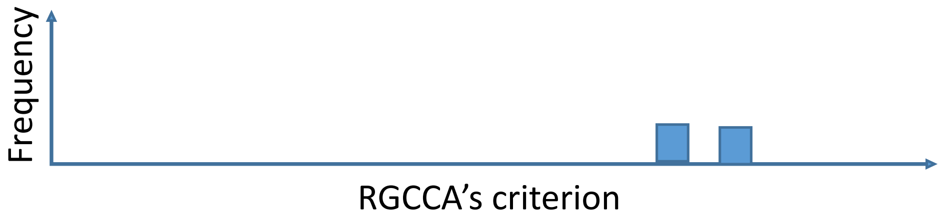
Permutation n°2

Parameter set n°1

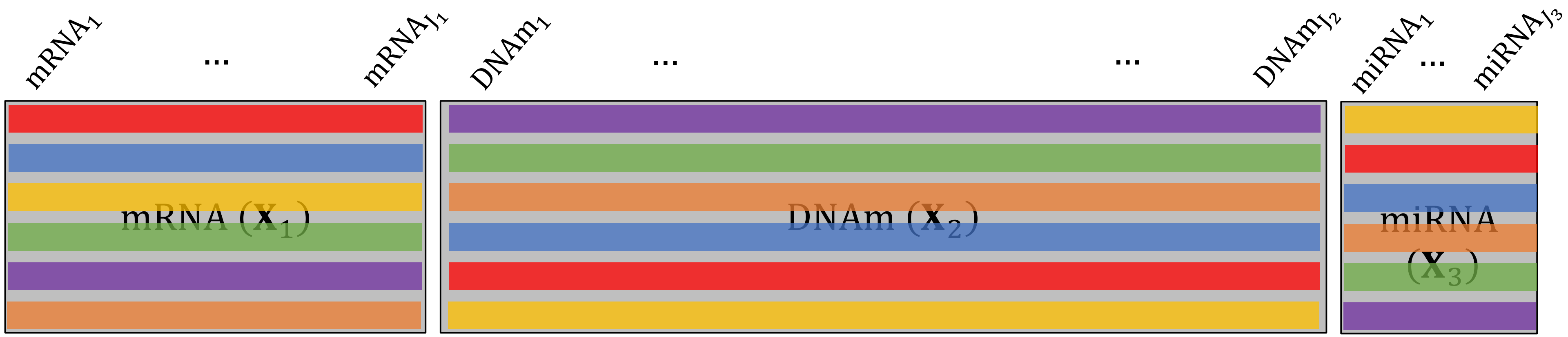


⋮ ⋮

Parameter set n°K

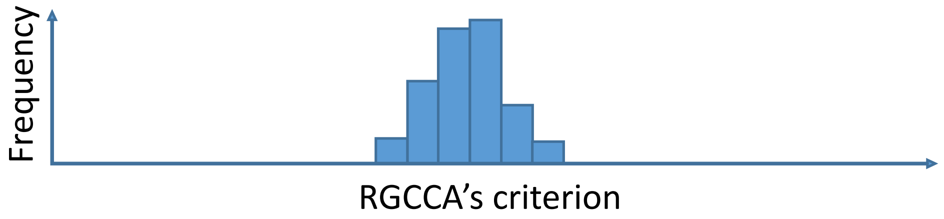


Tune parameters in an unsupervised setting



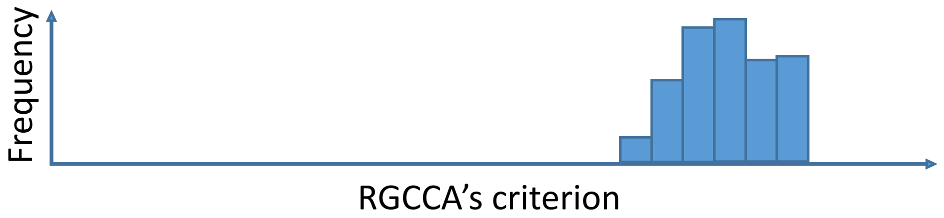
Permutation n°N

Parameter set n°1

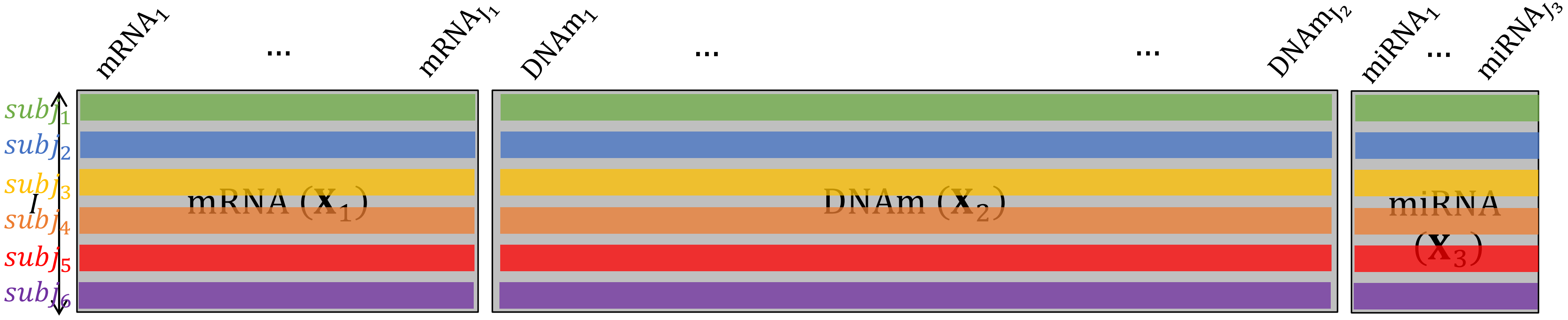


⋮ ⋮

Parameter set n°K

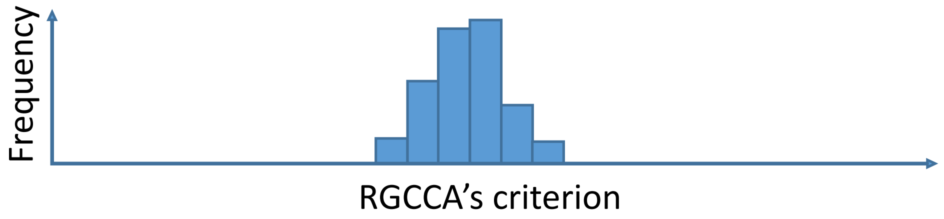


Tune parameters in an unsupervised setting



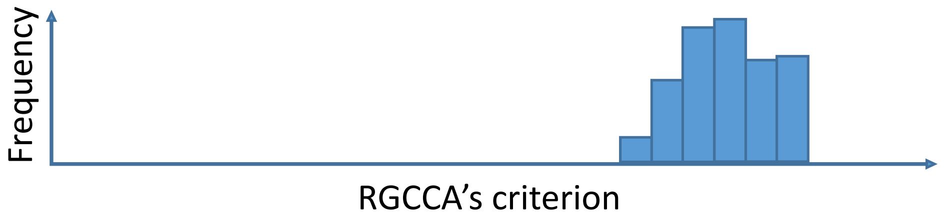
No Permutation

Parameter set n°1

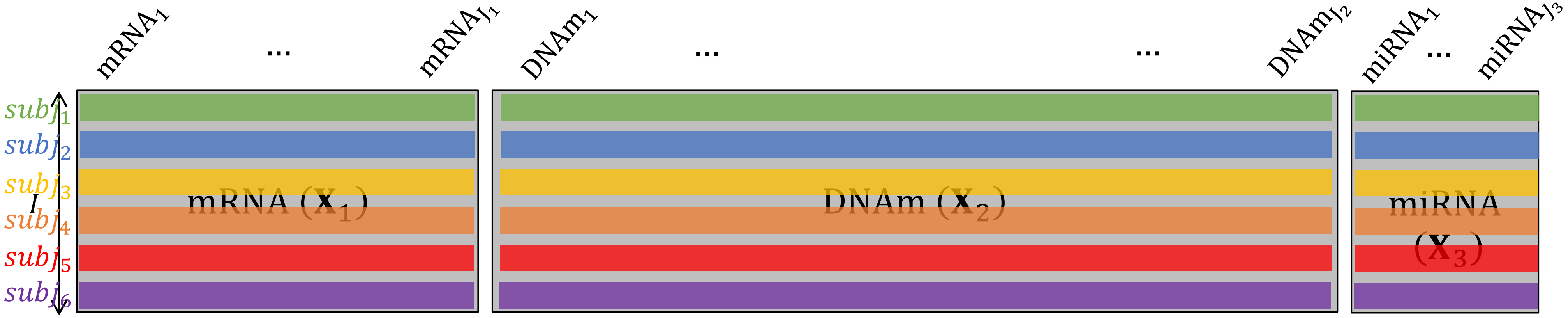


⋮

Parameter set n°K

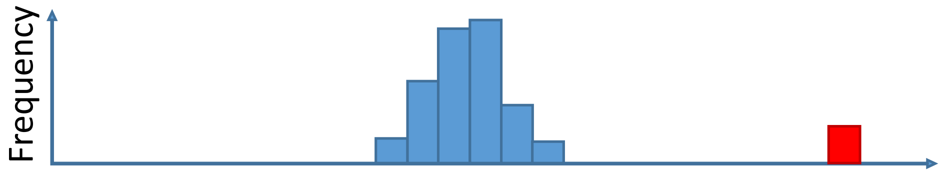


Tune parameters in an unsupervised setting



No Permutation

Parameter set n°1

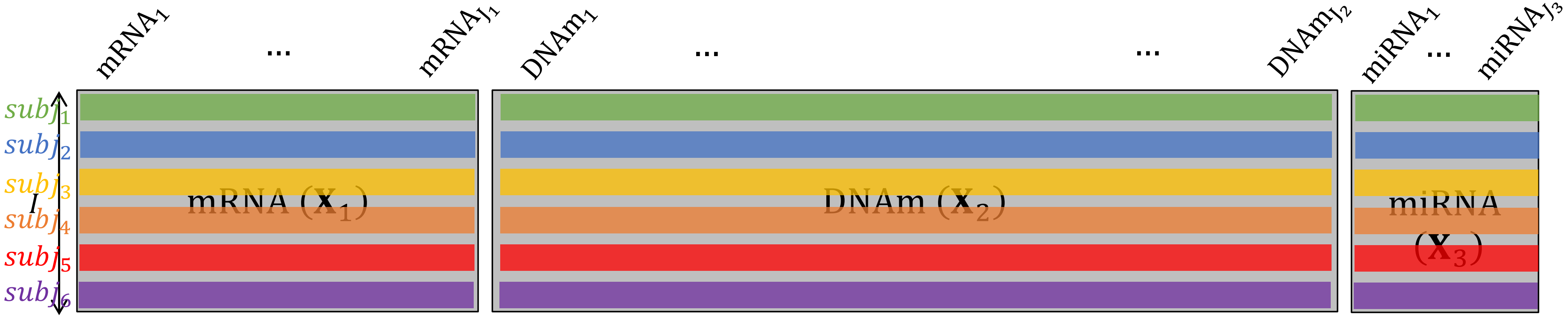


⋮

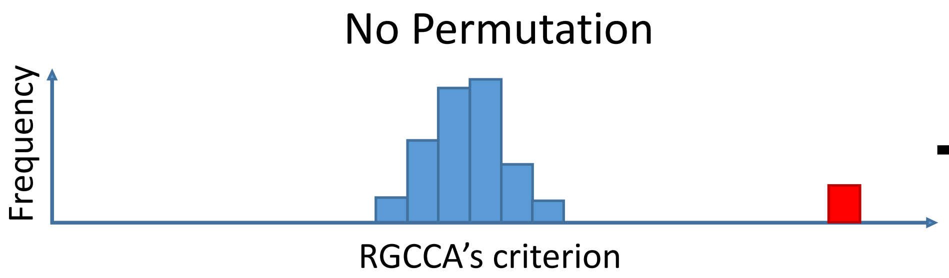
⋮

Parameter set n°K

Tune parameters in an unsupervised setting



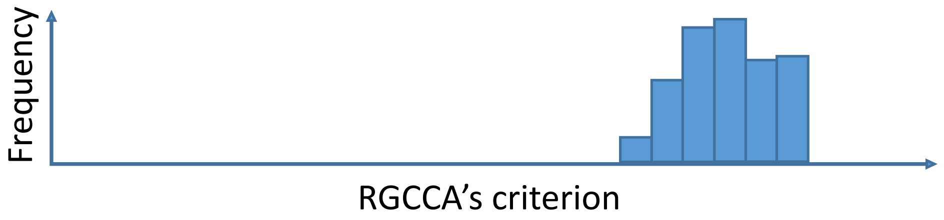
Parameter set n°1



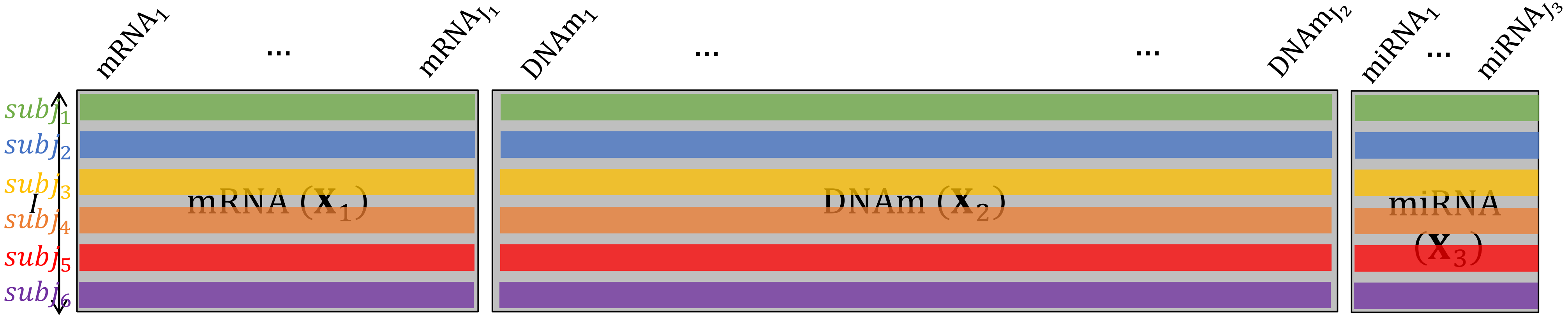
➔ The set of parameters is likely to be selected.

⋮

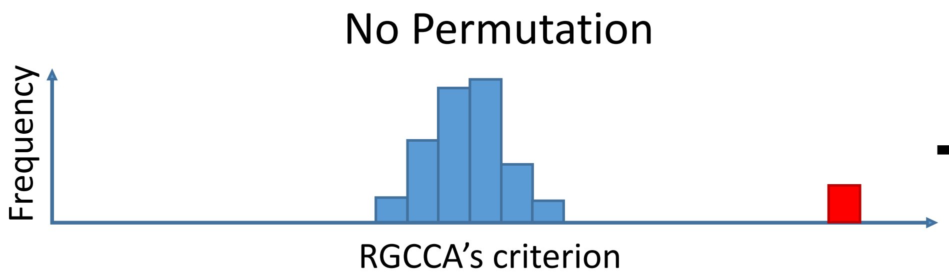
Parameter set n°K



Tune parameters in an unsupervised setting



Parameter set n°1

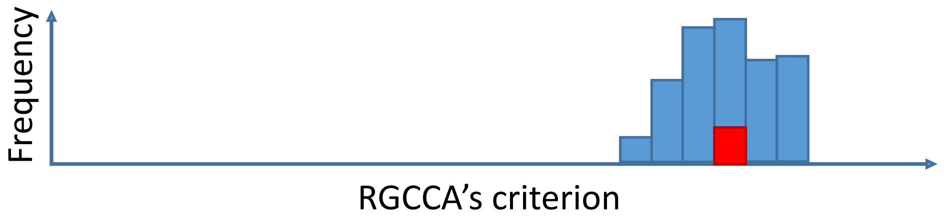


➔ The set of parameters is likely to be selected.

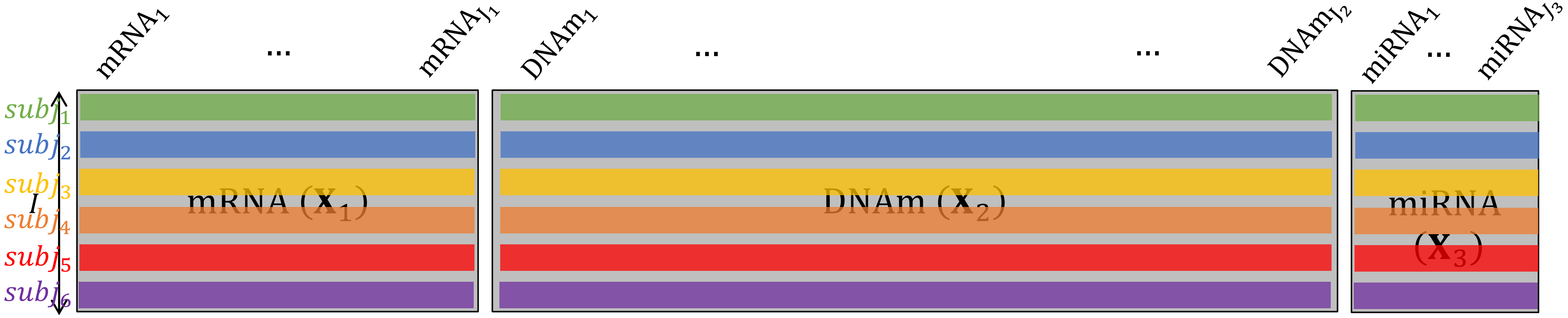
⋮

⋮

Parameter set n°K

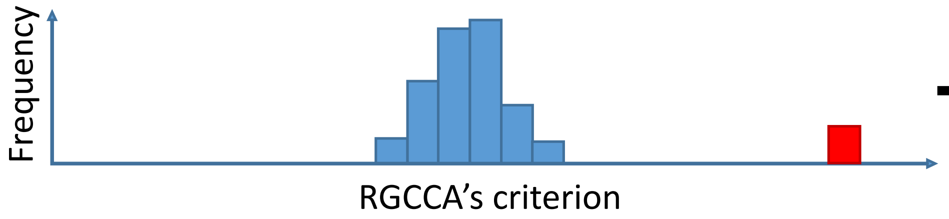


Tune parameters in an unsupervised setting



No Permutation

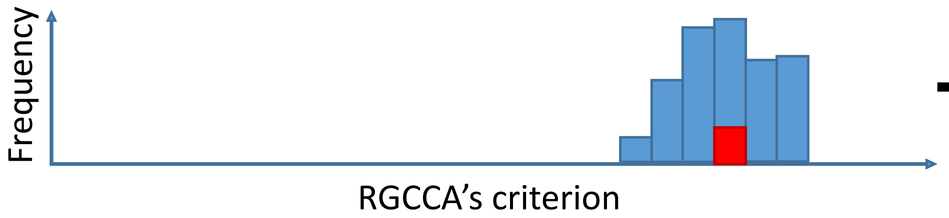
Parameter set n°1



➔ The set of parameters is likely to be selected.

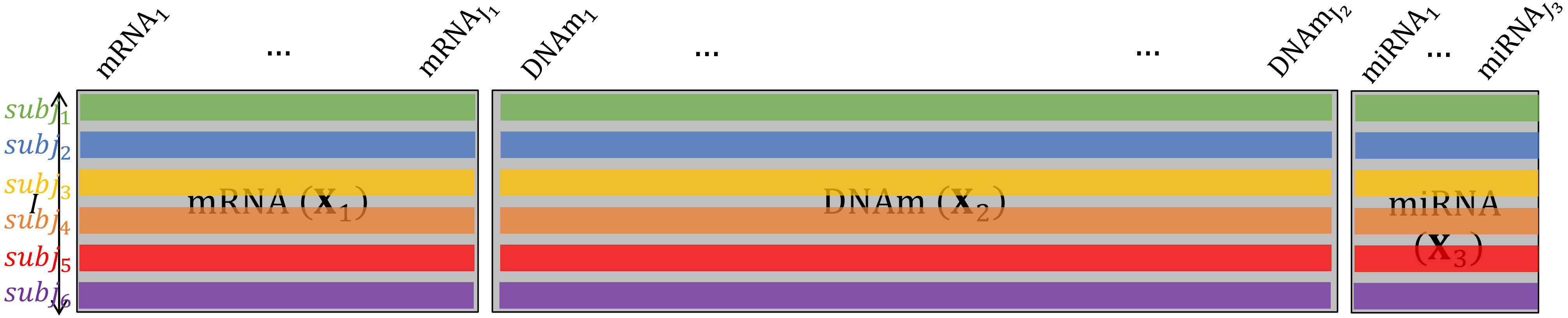
⋮ ⋮

Parameter set n°K

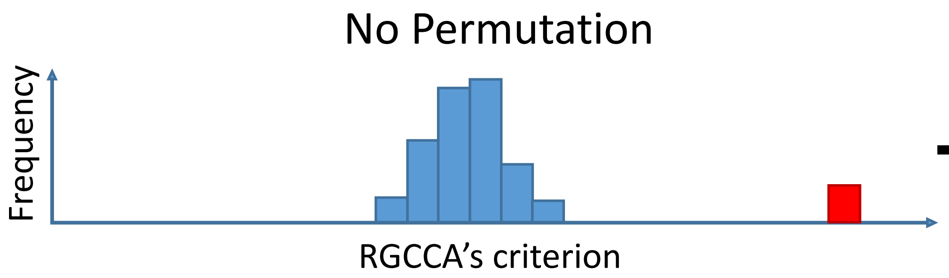


➔ The set of parameters is unlikely to be selected.

Tune parameters in an unsupervised setting



Parameter set n°1



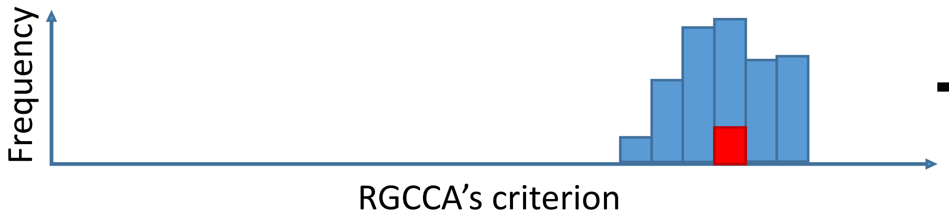
➔ The set of parameters is likely to be selected.

⋮

RGCCA choose the best set of parameters as the one with the highest value of $Z_k = \frac{(crit_{unperm} - \mu_{crit}^{perm})}{\sigma_{crit}^{perm}}$

⋮

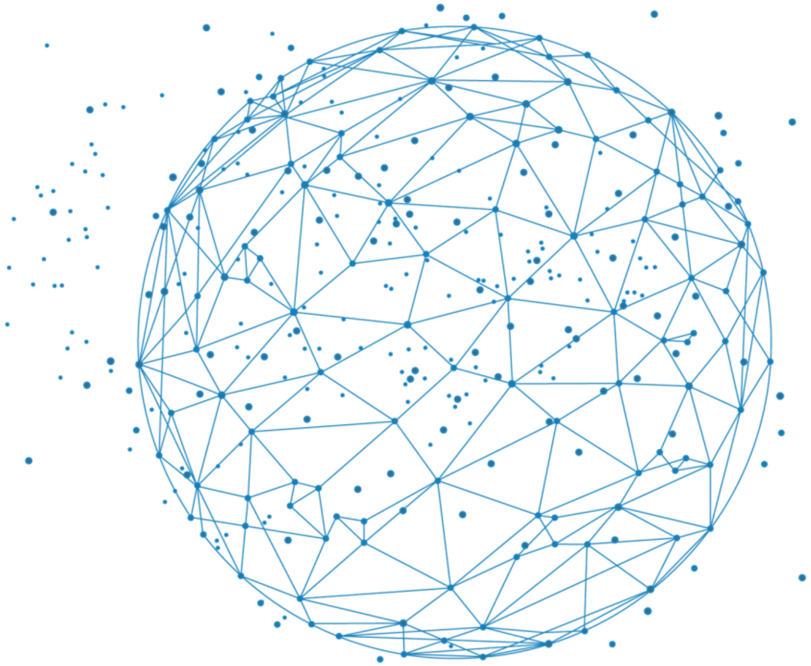
Parameter set n°K



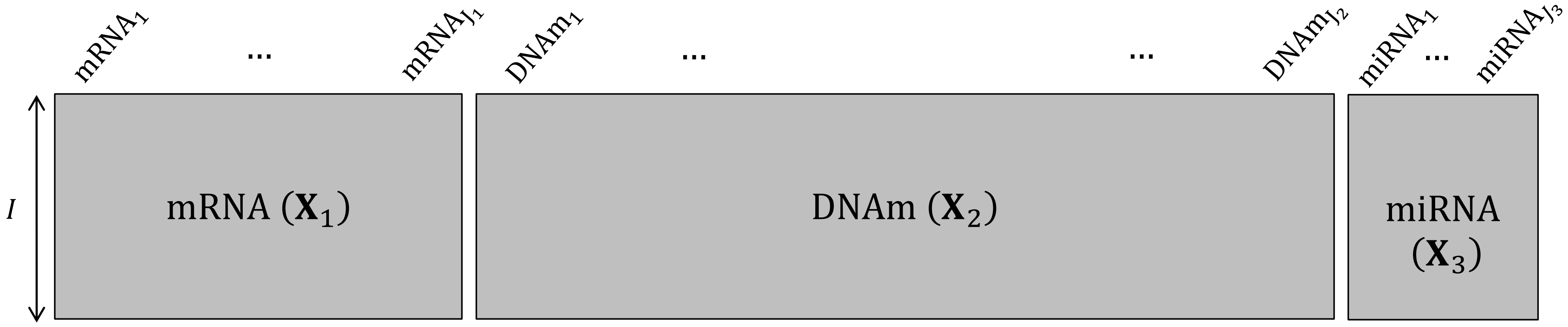
➔ The set of parameters is unlikely to be selected.

Let us apply this permutation procedure on the MDD case study

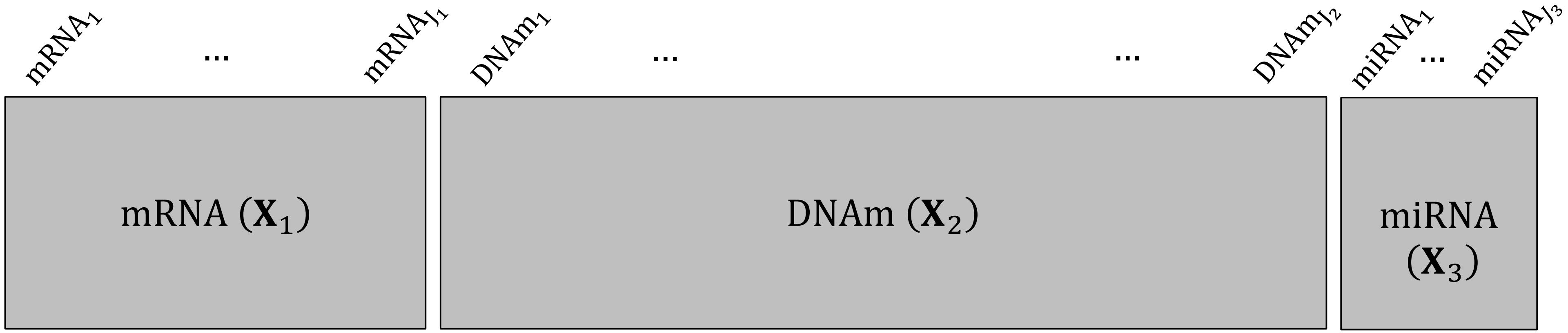
➔ See section 3.3 on the Rmarkdown `MDD_case_study_RGCCA`



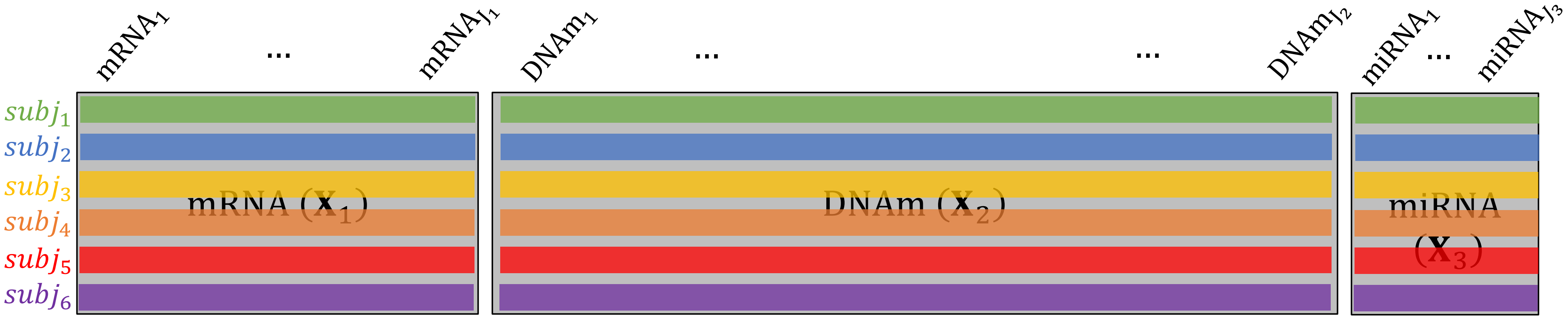
Evaluate the robustness of the model by bootstrapping.



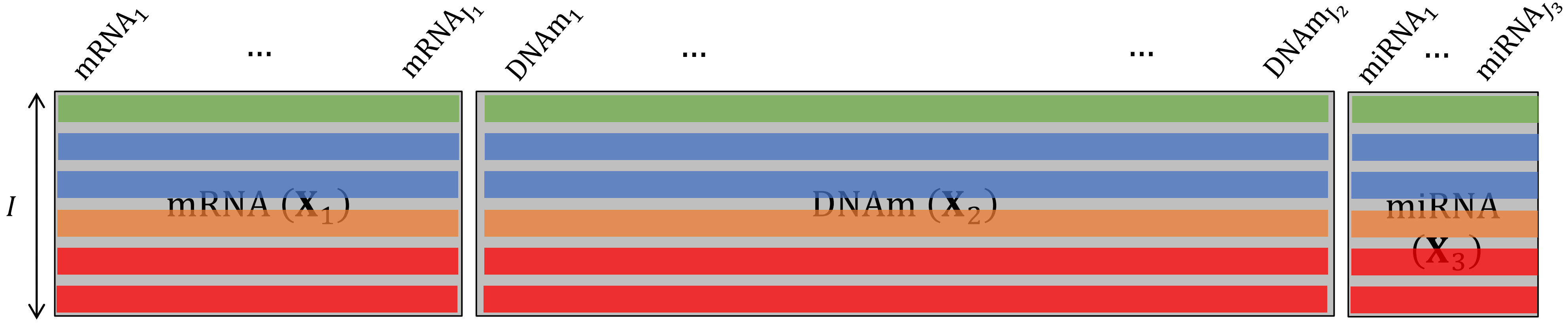
Evaluate the robustness of the model by bootstrapping.



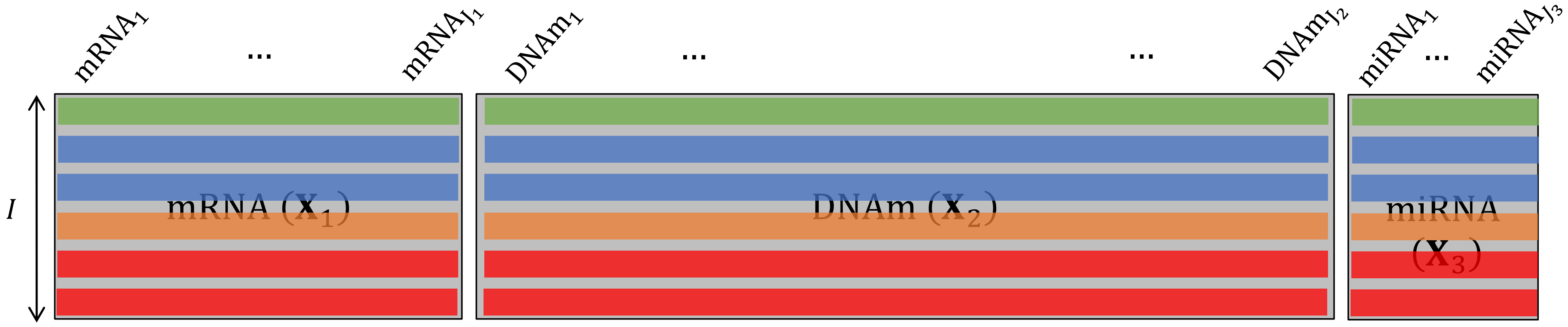
Evaluate the robustness of the model by bootstrapping.



Evaluate the robustness of the model by bootstrapping.

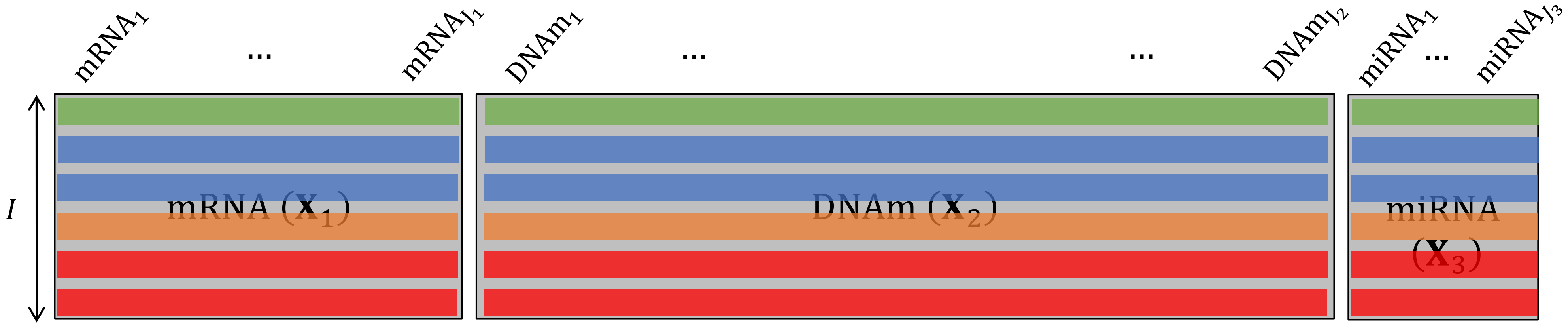


Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°1

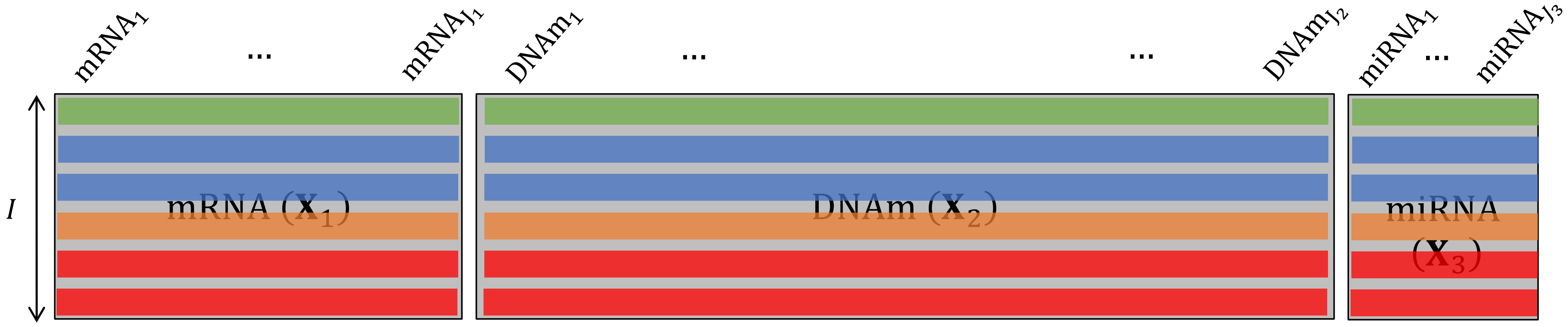
Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°1

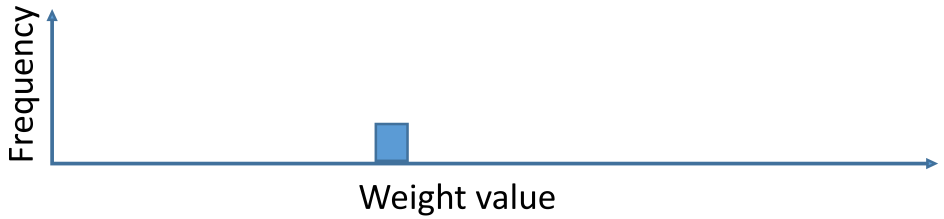
Weight for mRNA_1

Evaluate the robustness of the model by bootstrapping.

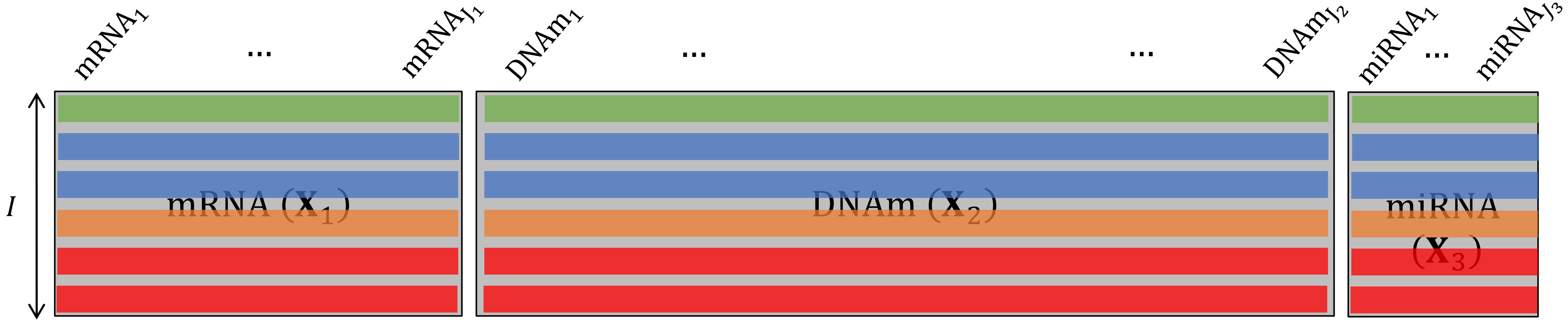


Bootstrap sample n°1

Weight for mRNA_1

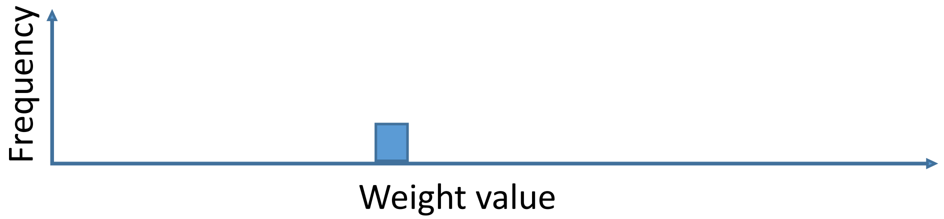


Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°1

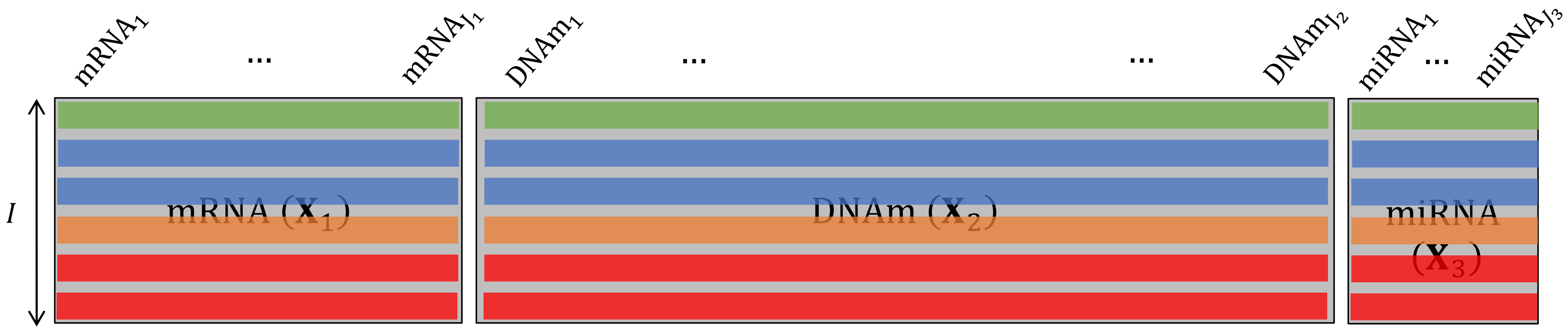
Weight for $mRNA_1$



⋮

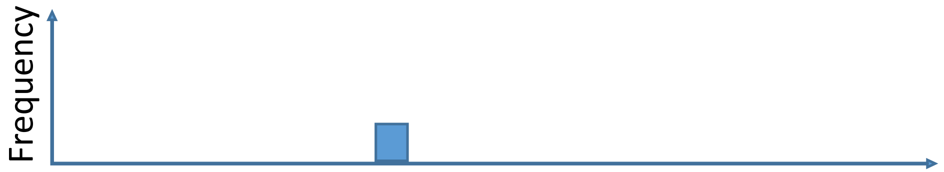
Weight for $miRNA_{J_3}$

Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°1

Weight for $mRNA_1$

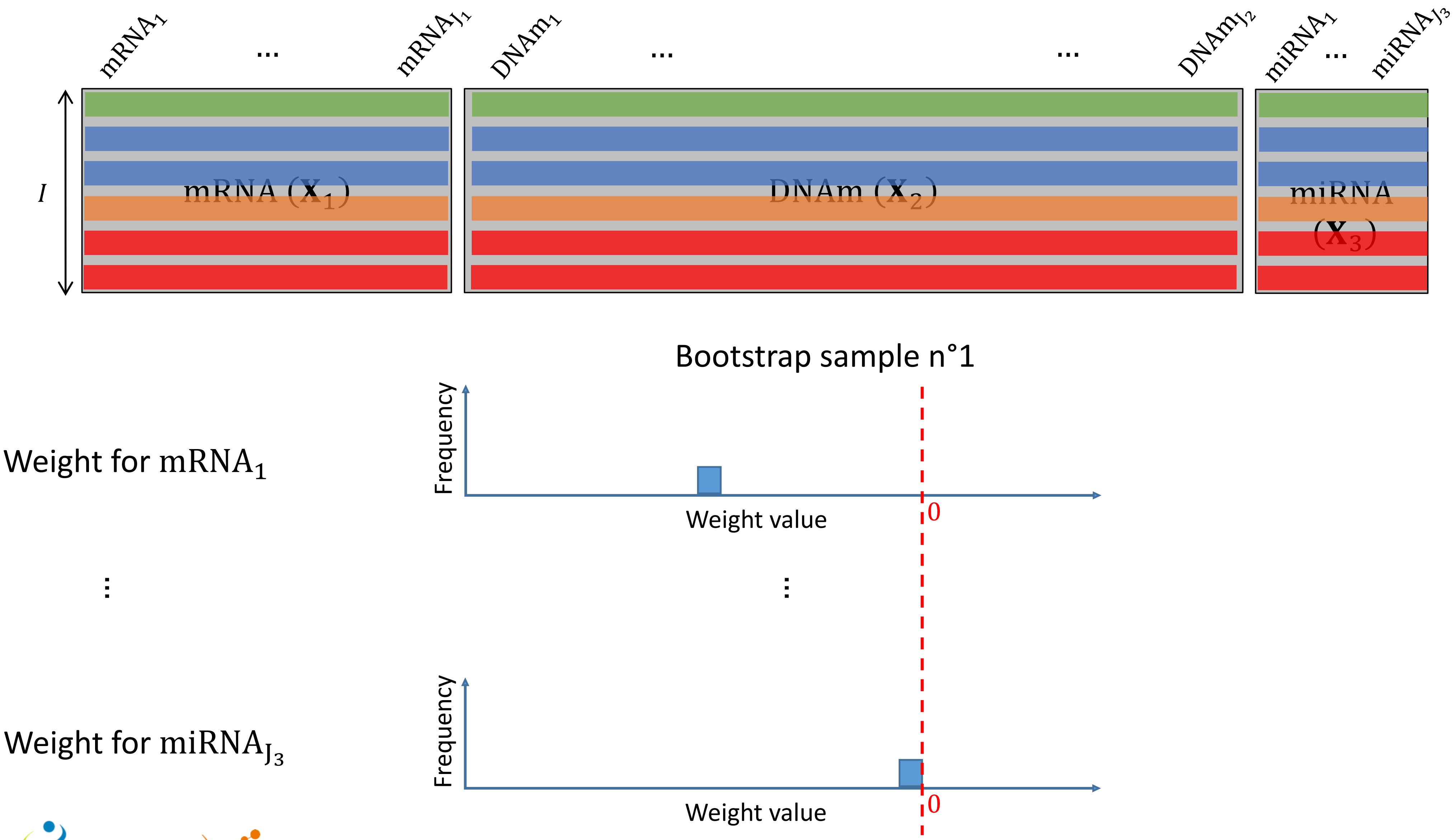


⋮

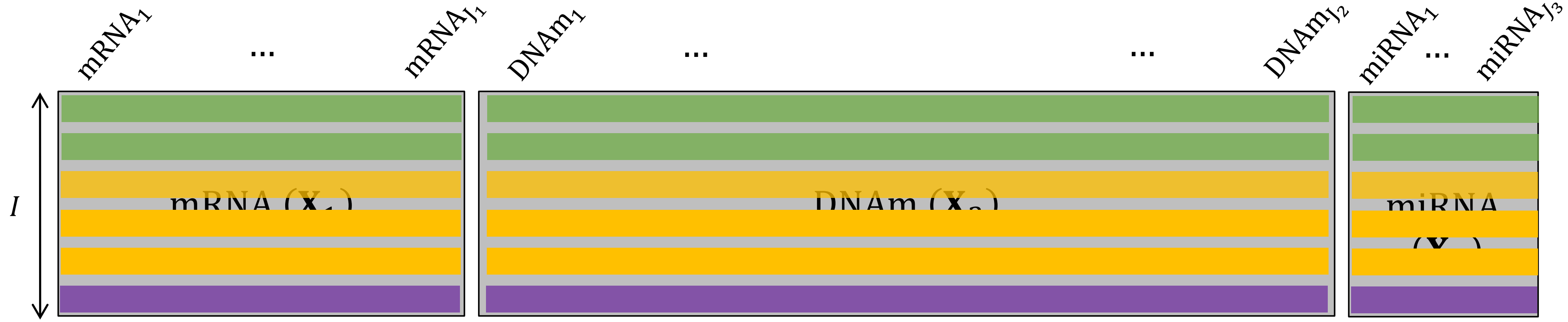
⋮

Weight for $miRNA_{J_3}$

Evaluate the robustness of the model by bootstrapping.



Evaluate the robustness of the model by bootstrapping.

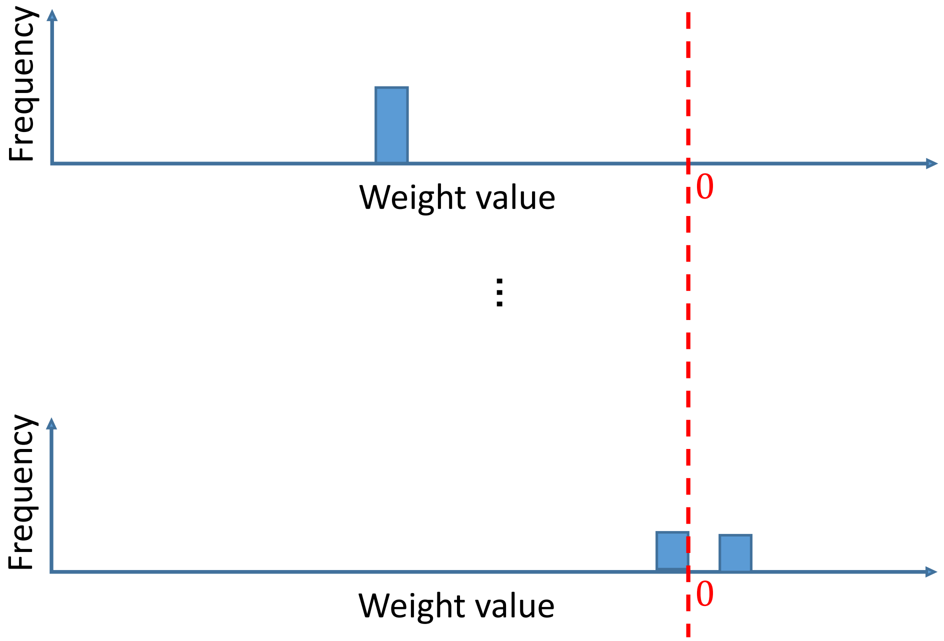


Bootstrap sample n°2

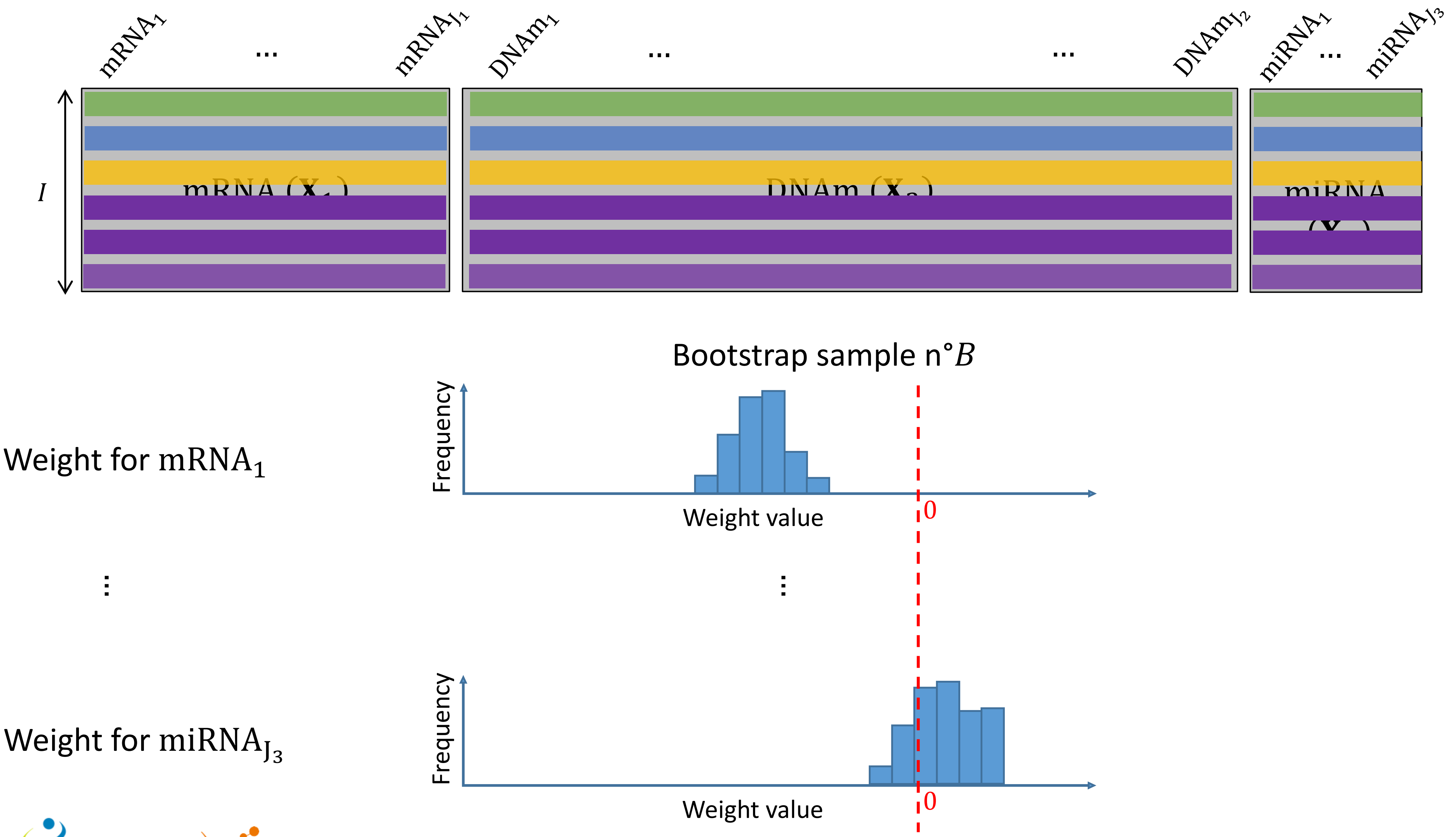
Weight for $mRNA_1$

⋮

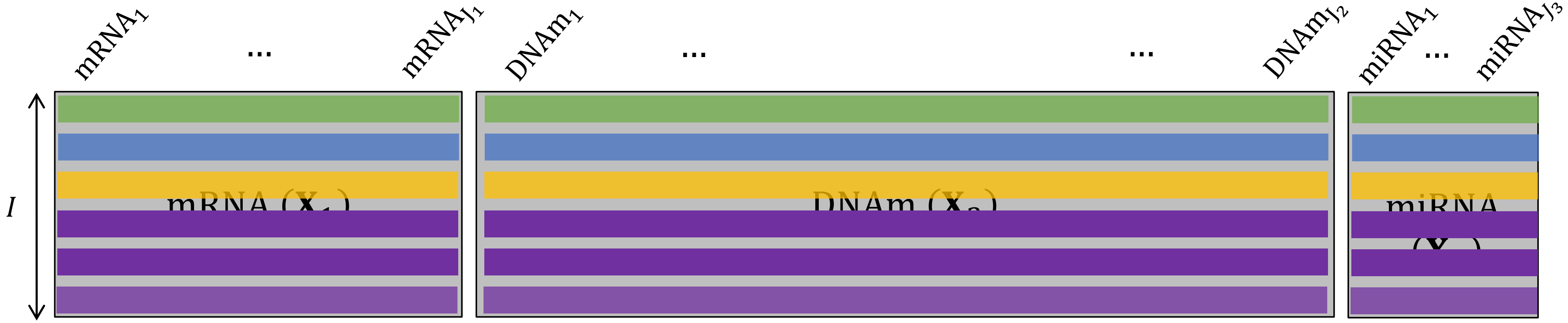
Weight for $miRNA_{J_3}$



Evaluate the robustness of the model by bootstrapping.



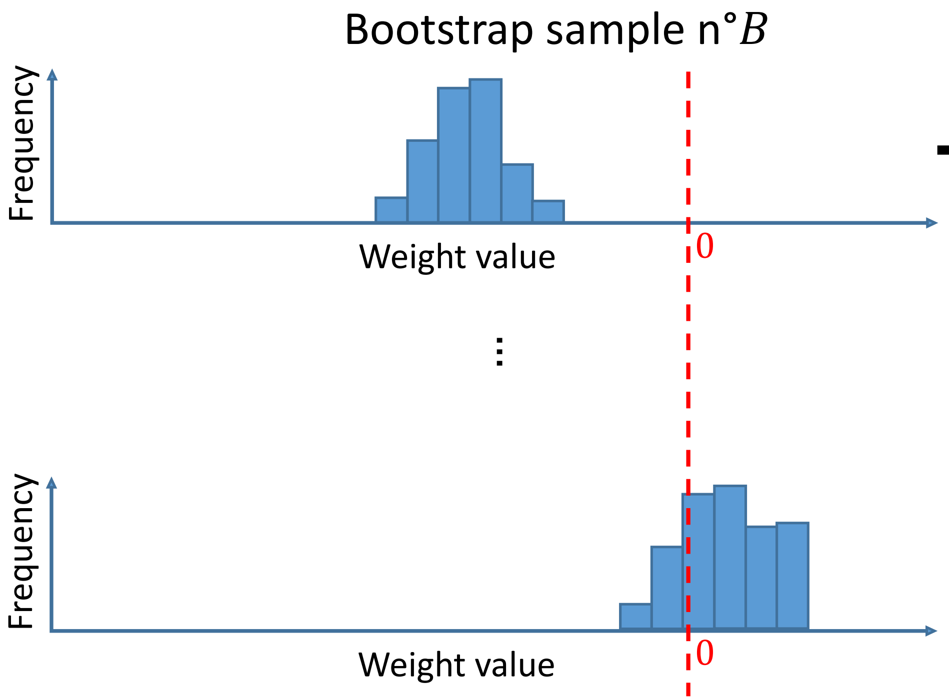
Evaluate the robustness of the model by bootstrapping.



Weight for mRNA_1

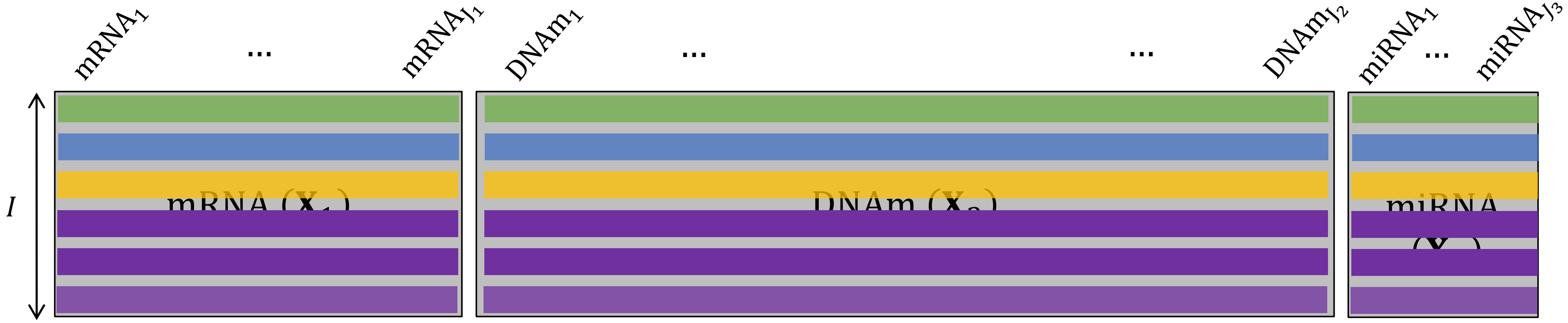
⋮

Weight for miRNA_{J_3}



➔ The weight is likely to be considered as It is significantly different from 0.

Evaluate the robustness of the model by bootstrapping.

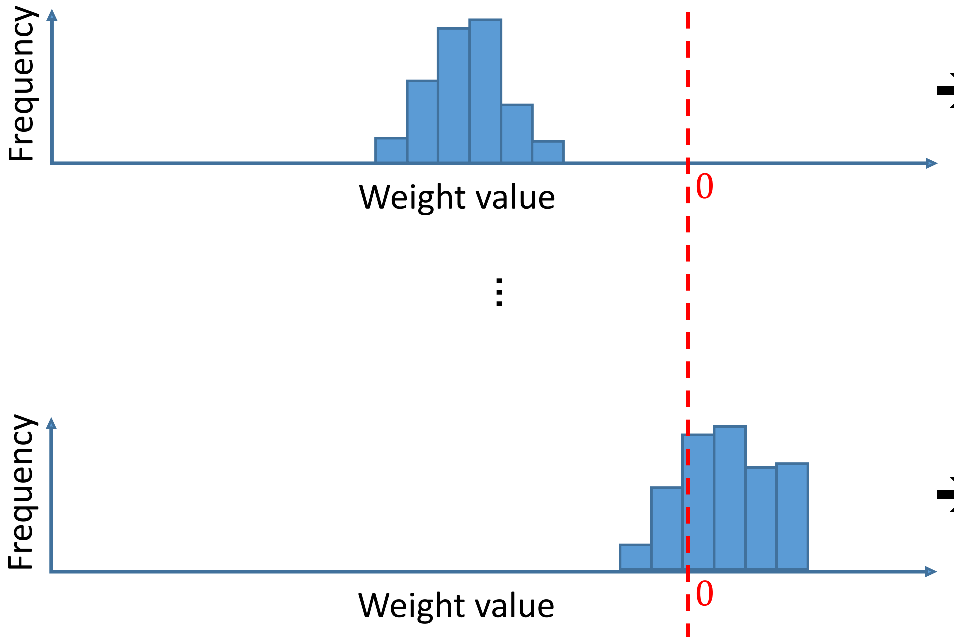


Bootstrap sample n°B

Weight for $mRNA_1$

⋮

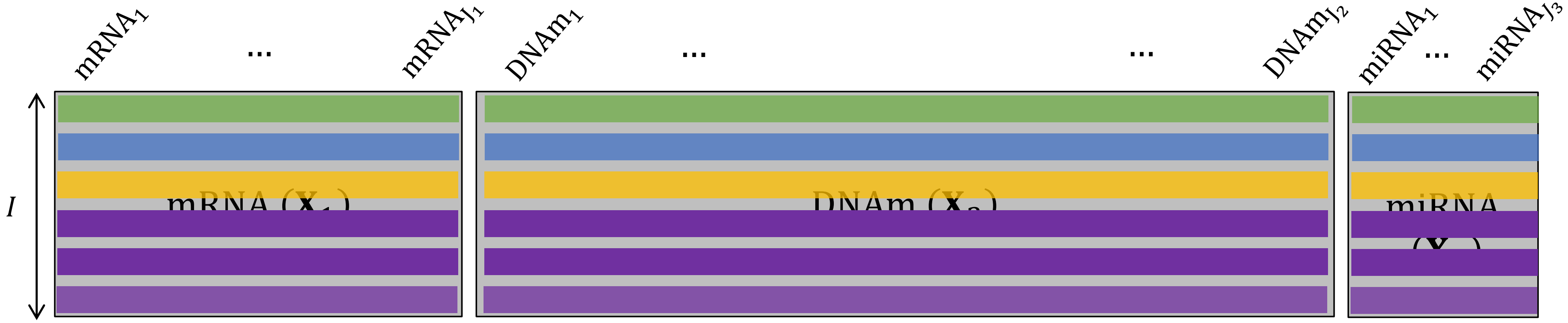
Weight for $miRNA_{J_3}$



➔ The weight is likely to be considered as
It is significantly different from 0.

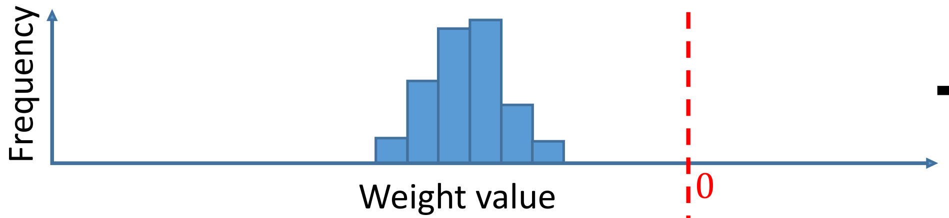
➔ The weight is unlikely to be considered as
It is not significantly different from 0.

Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°B

Weight for $mRNA_1$



➔ The weight is likely to be considered as It is significantly different from 0.

⋮

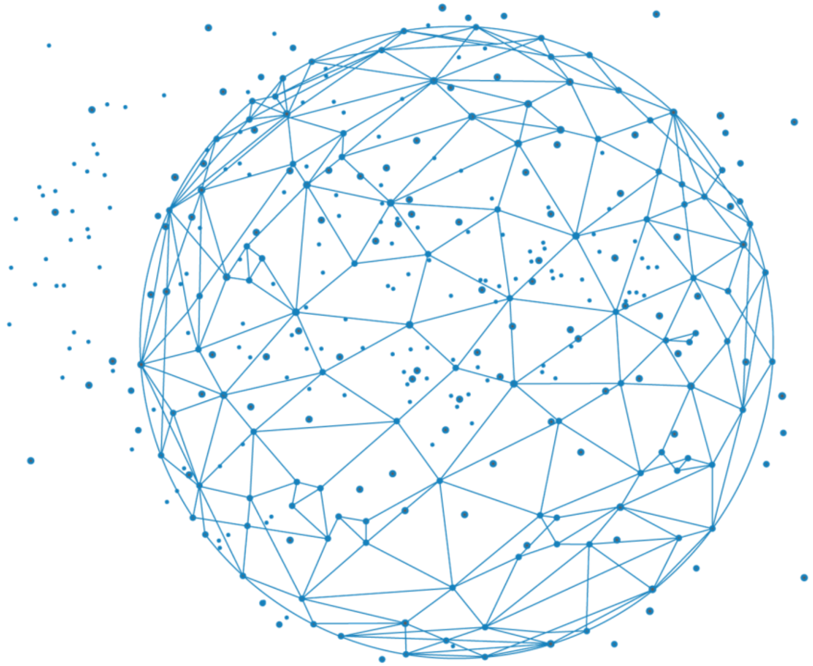
Out of these distributions, RGCCA non-parametrically estimates confidence intervals $([q_{0.025}, q_{0.975}])$ and p-values $\left(min(Nb_{\geq 0}, Nb_{\leq 0})/max(Nb_{\geq 0}, Nb_{\leq 0})\right)$.

Weight for $miRNA_{J_3}$

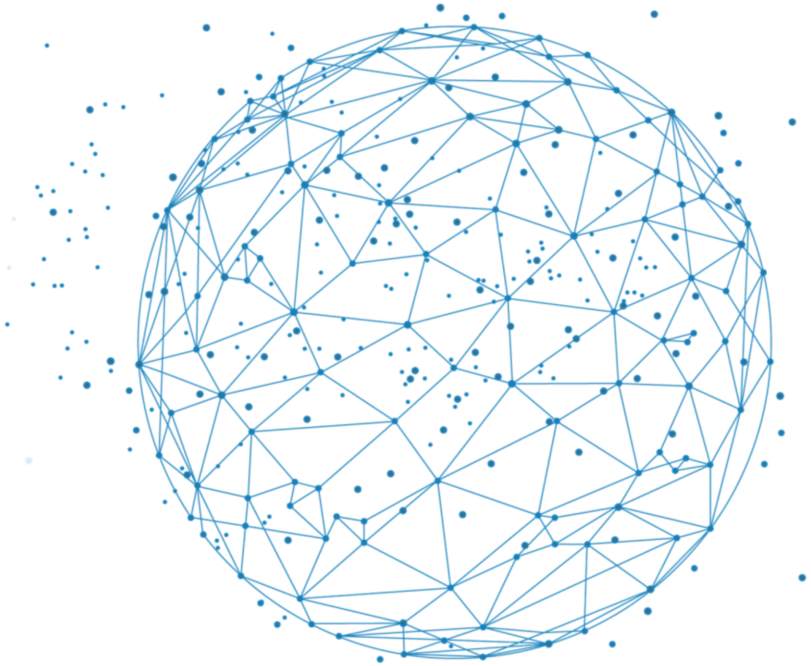
➔ The weight is unlikely to be considered as It is not significantly different from 0.

Let us apply this permutation procedure on the MDD case study

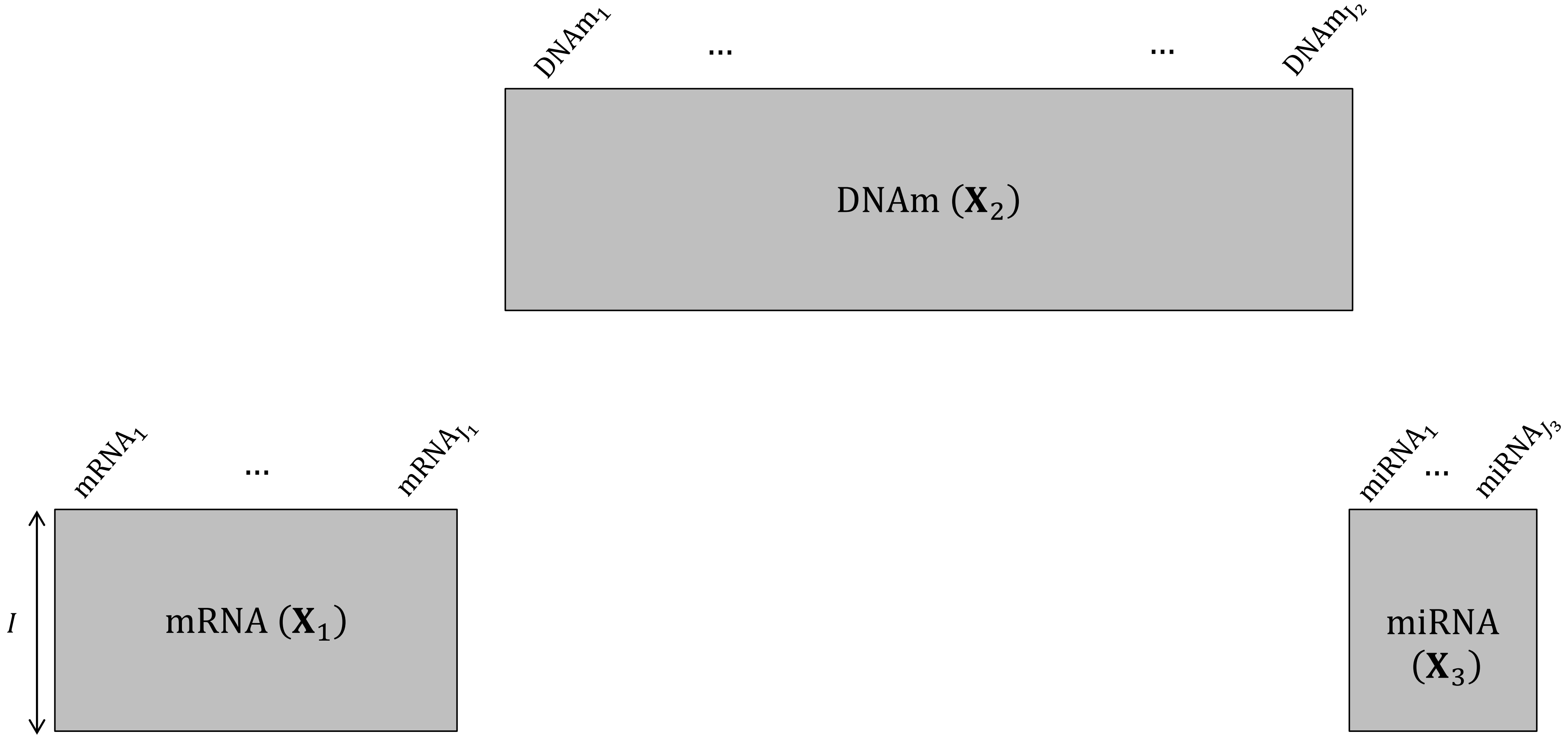
➔ See section 3.4 on the Rmarkdown `MDD_case_study_RGCCA`



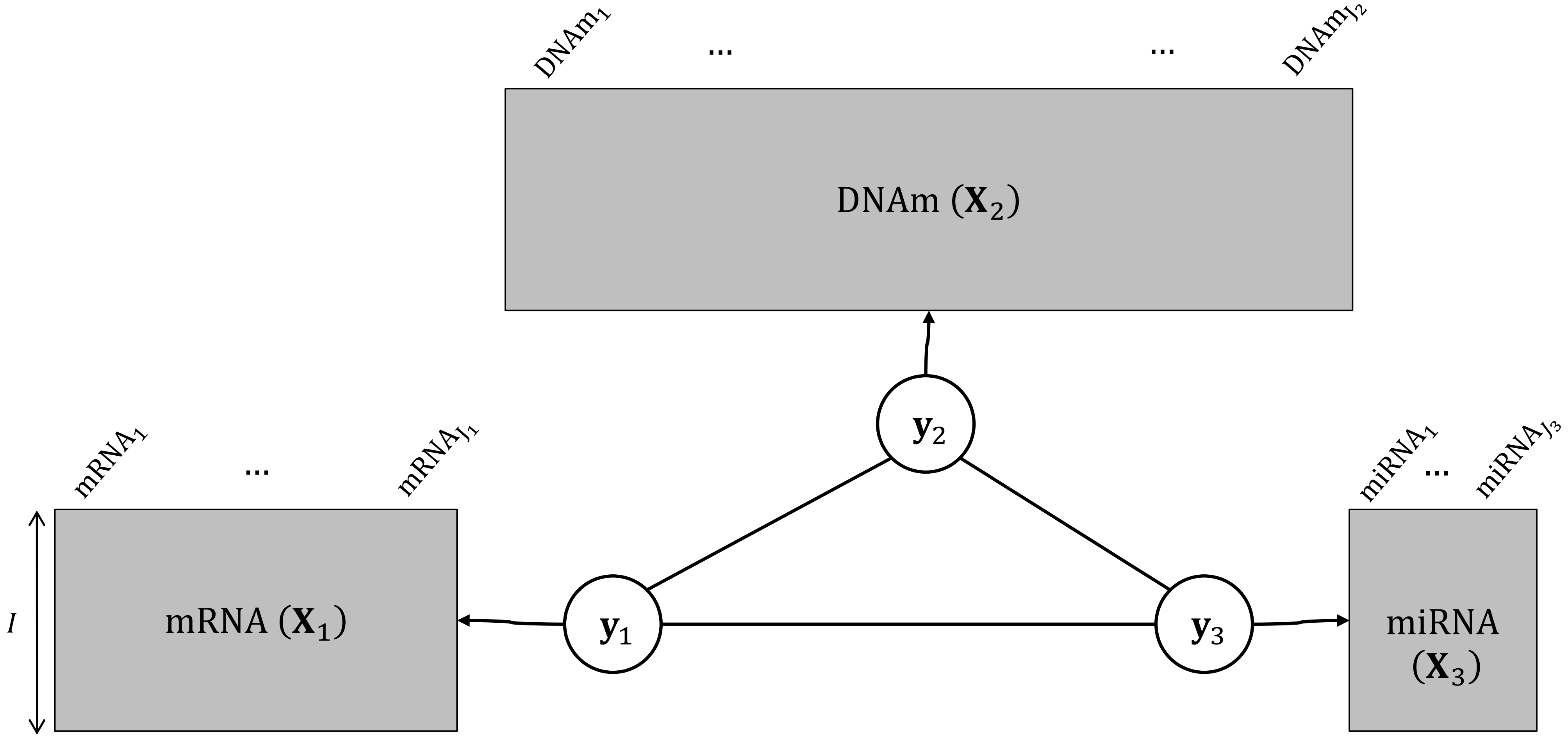
1. Introduction of the case study
2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)
3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)
4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)
5. **Supervised analysis with RGCCA**
6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)
7. **The flexible Optimization Framework of RGCCA**
   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global



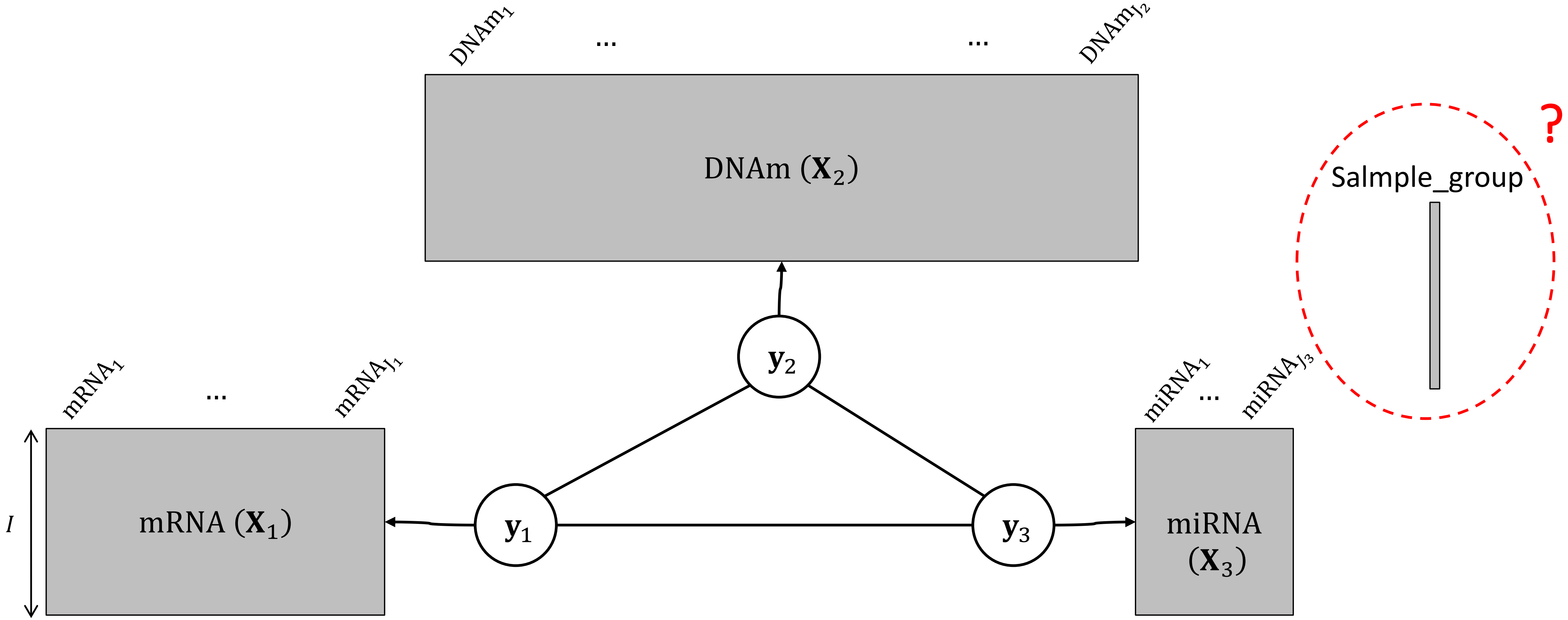
Supervising with RGCCA



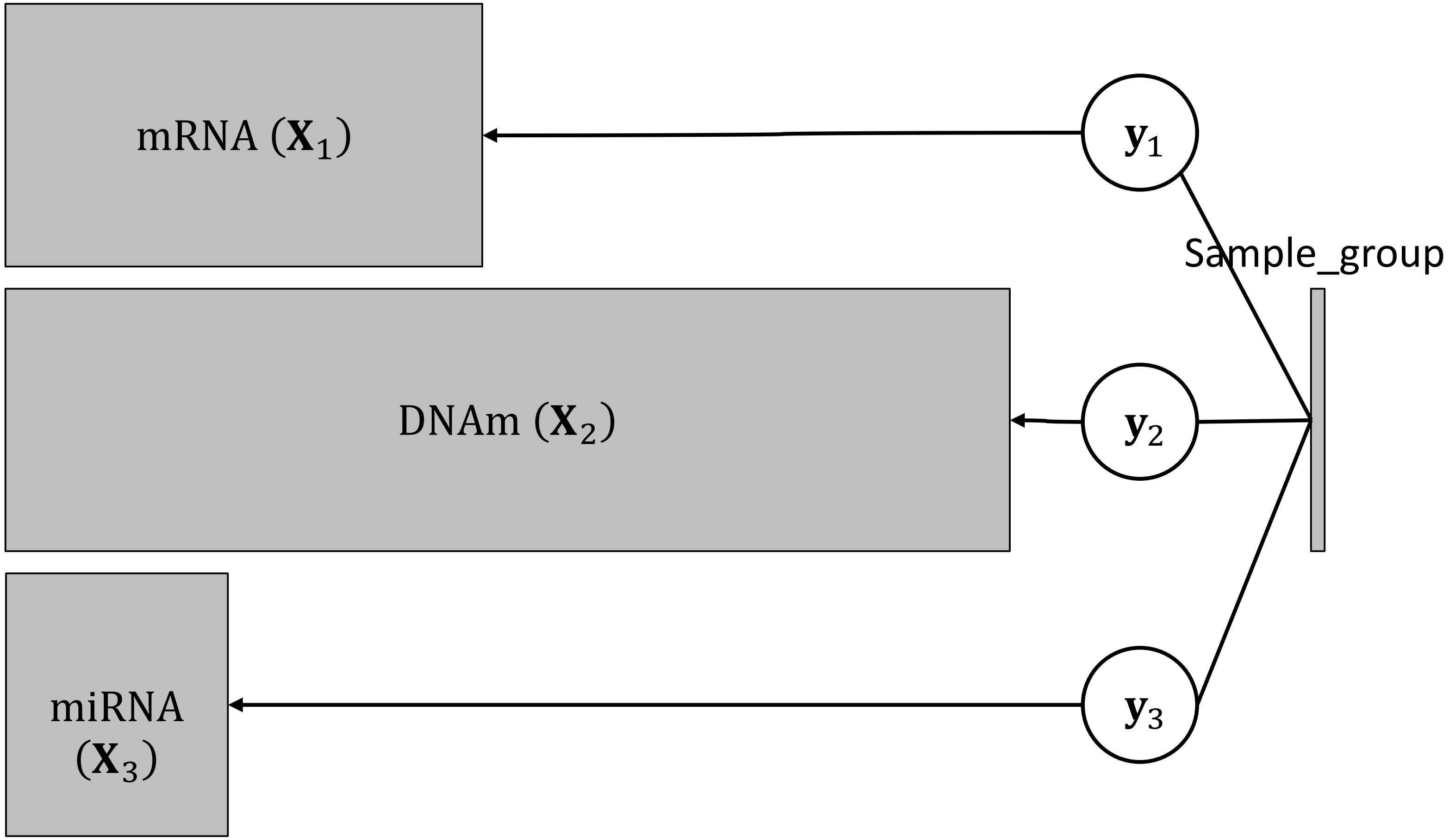
Supervising with RGCCA



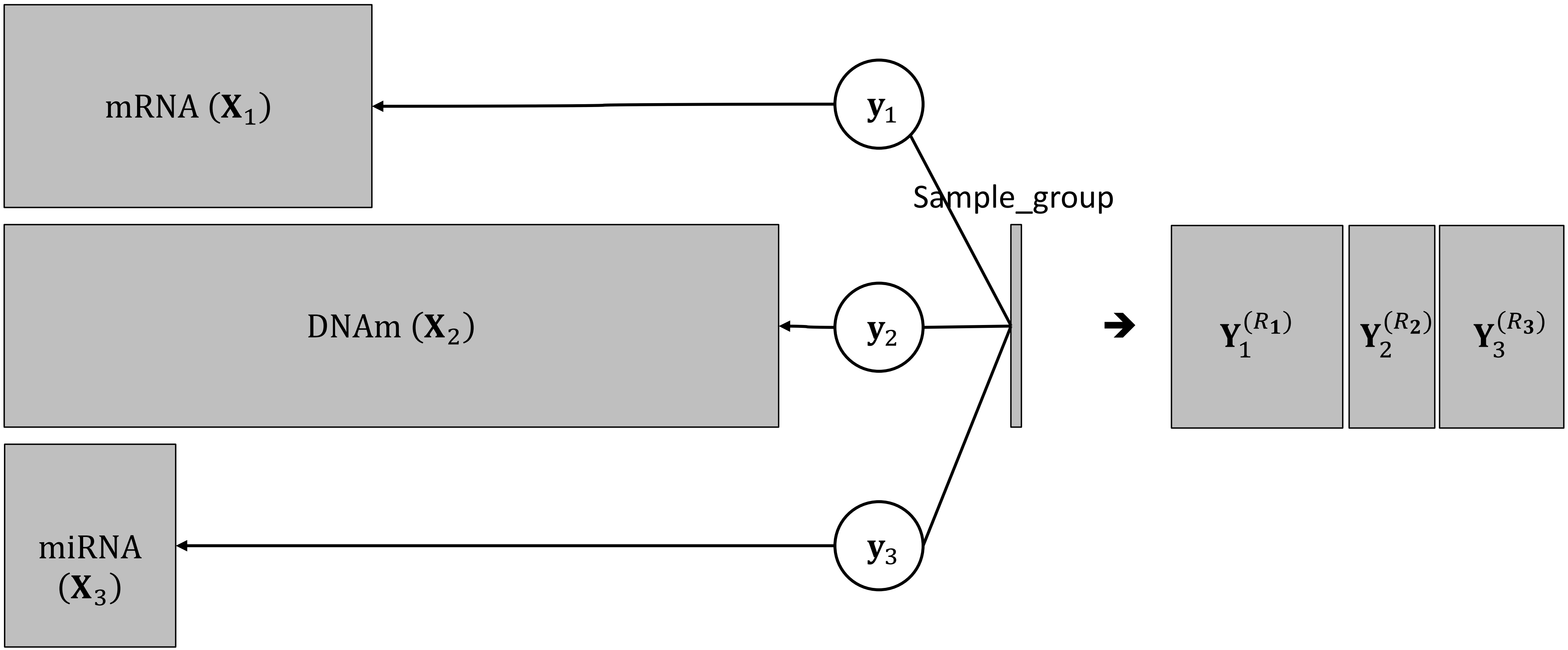
Supervising with RGCCA



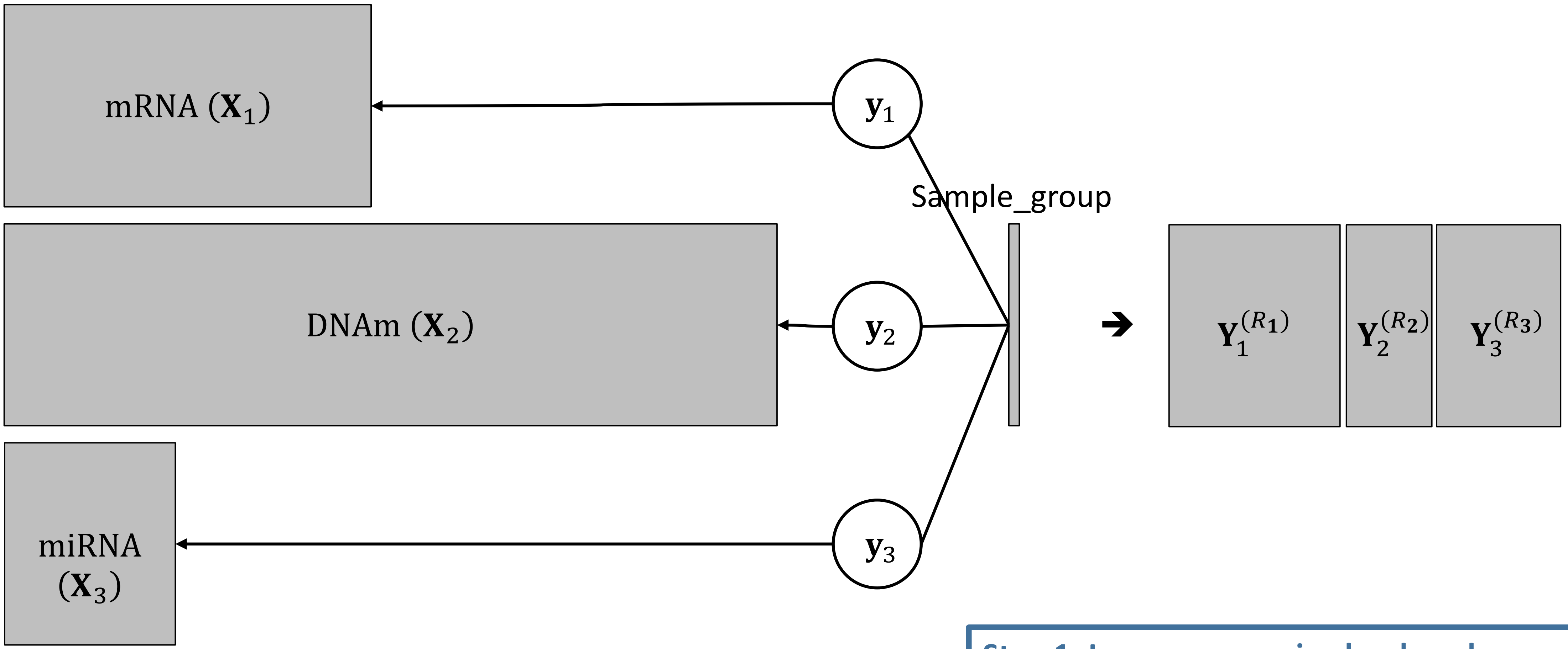
Supervising with RGCCA



Supervising with RGCCA

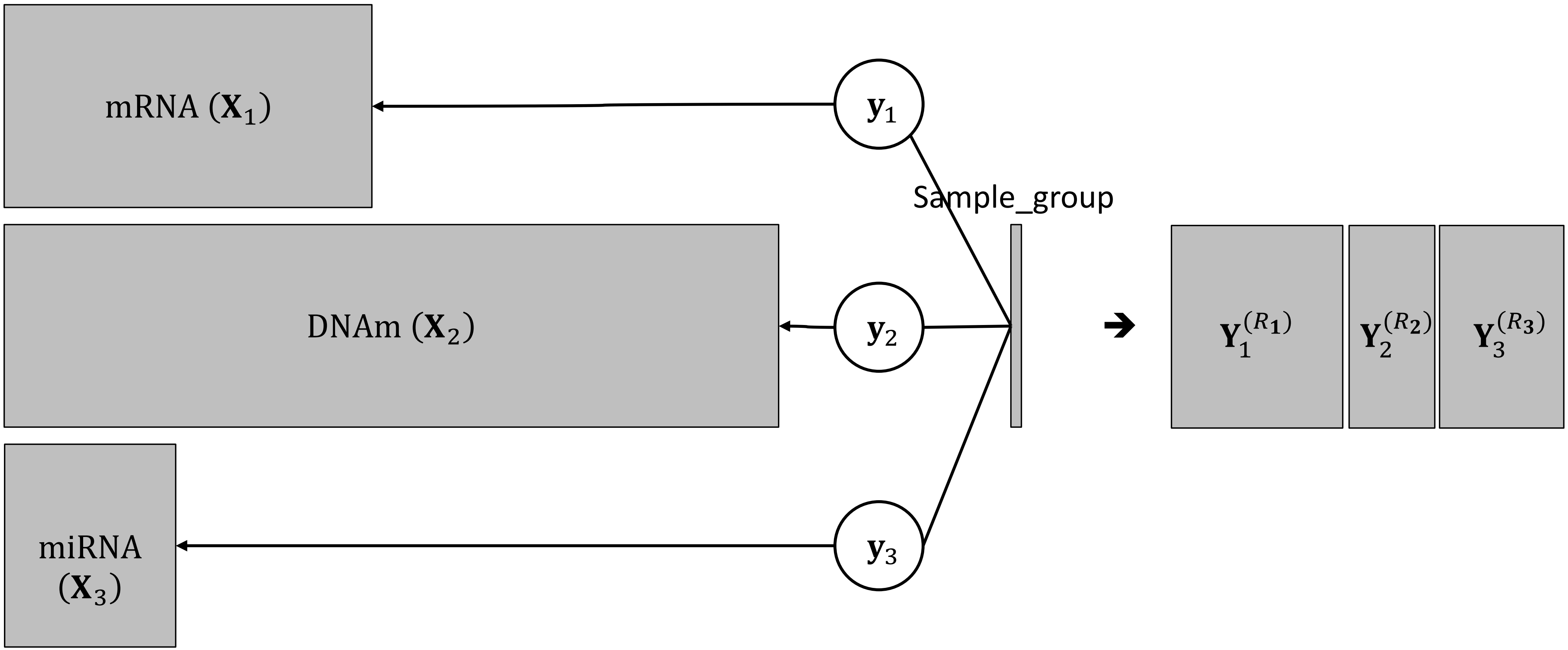


Supervising with RGCCA

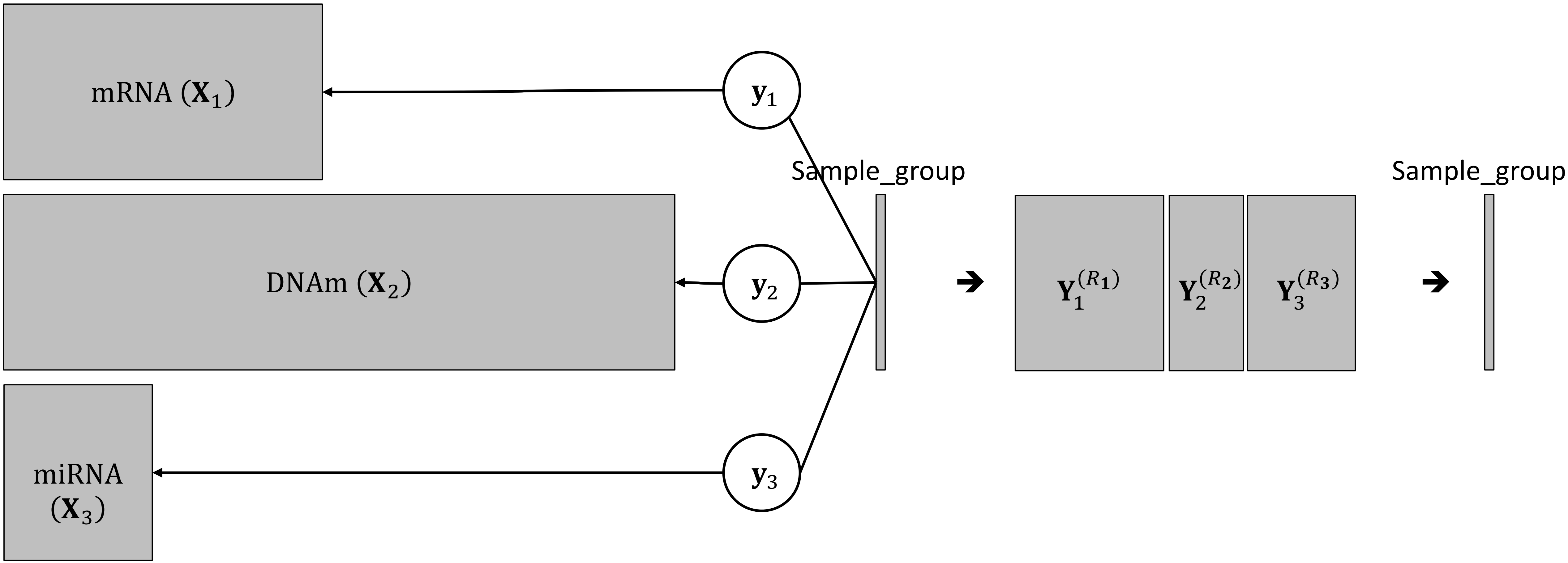


Step 1: Learn a supervised reduced space

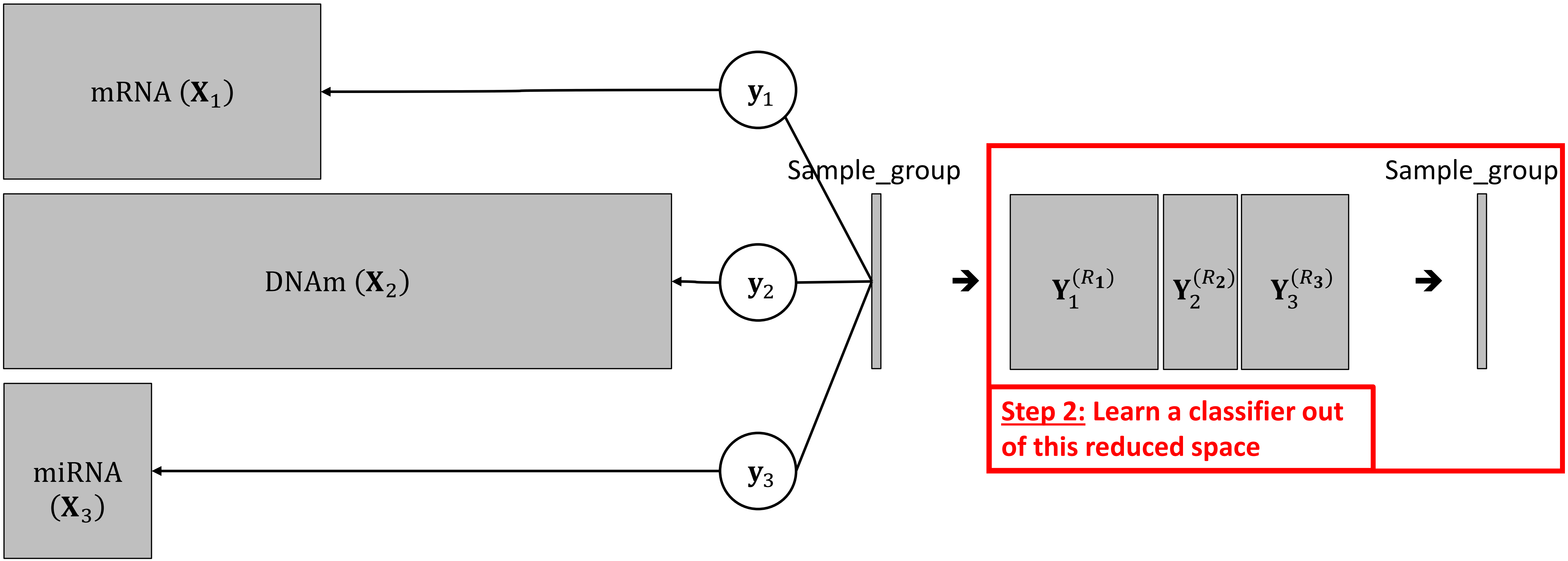
Supervising with RGCCA



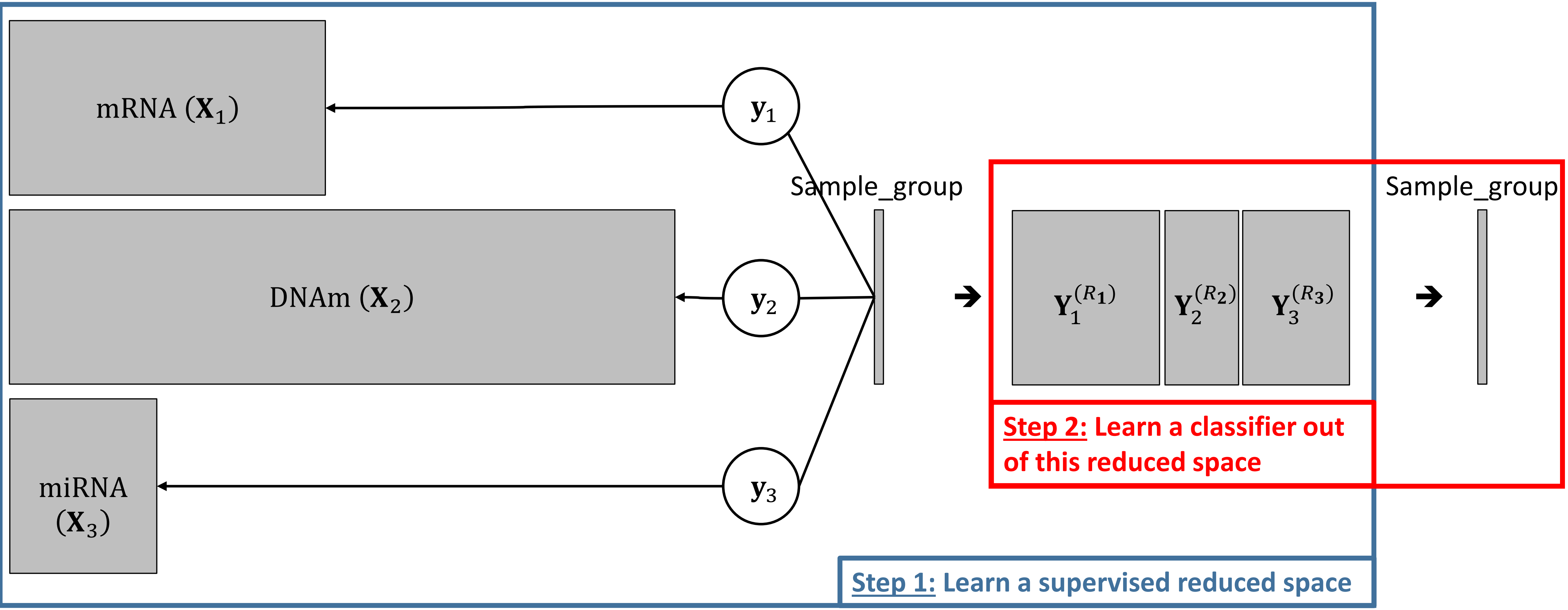
Supervising with RGCCA



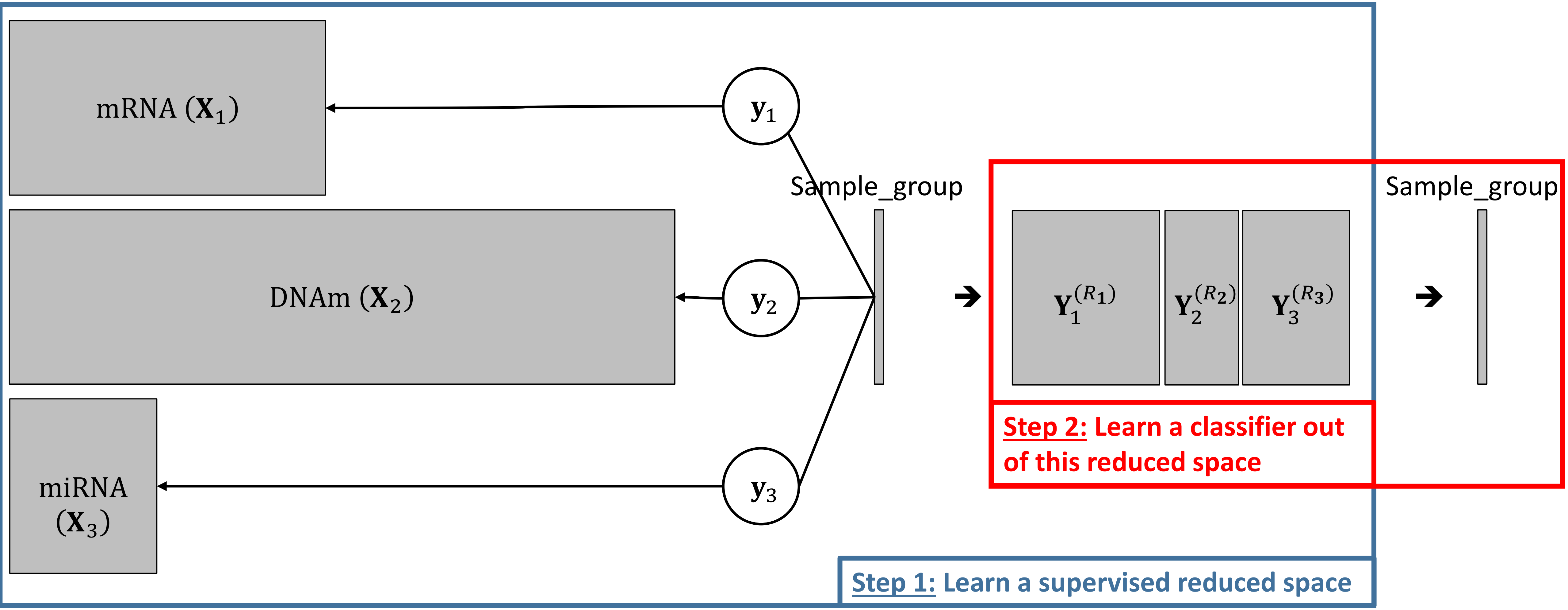
Supervising with RGCCA



Supervising with RGCCA

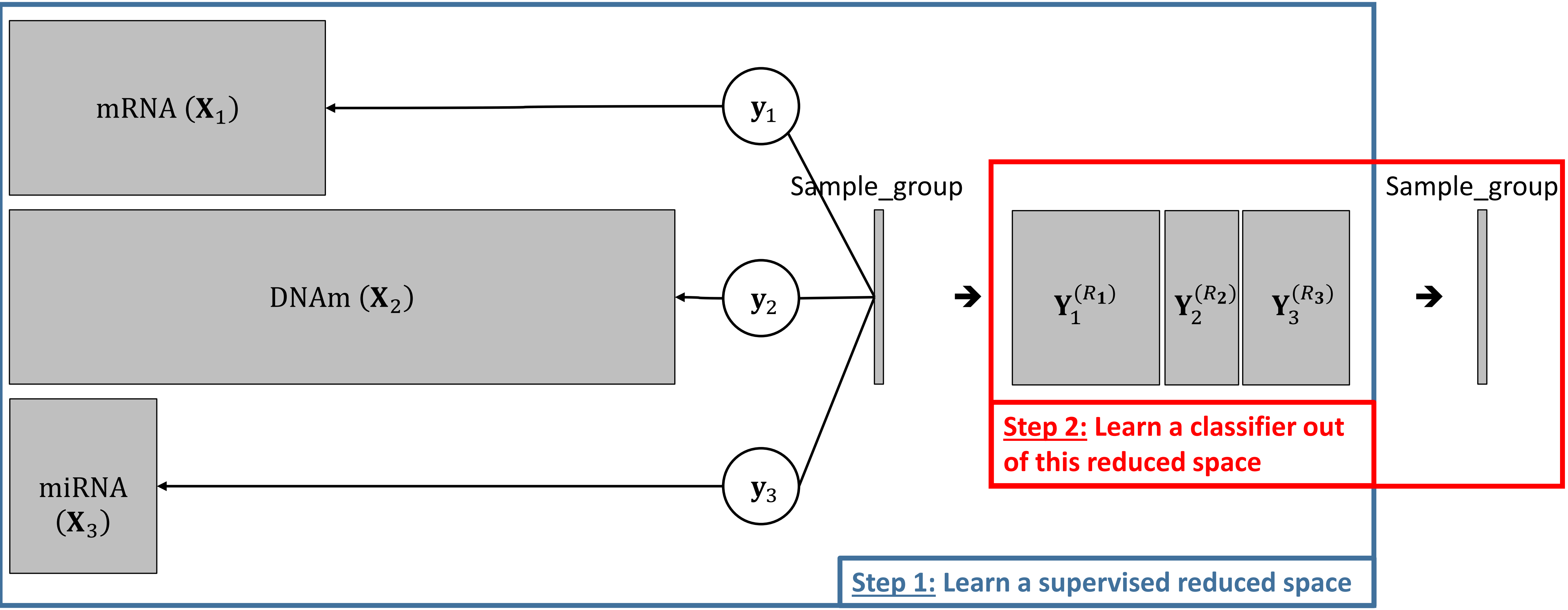


Supervising with RGCCA



The model sequentially learn block-weight vectors to compute components and a classifier.

Supervising with RGCCA



The model sequentially learn block-weight vectors to compute components and a classifier.

Standard Cross-Validation can be performed.

F1-score

F1-score

Confusion Matrix:

		True labels	
		Positive	Negative
Predicted labels	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

F1-score

Confusion Matrix:

		True labels	
		Positive	Negative
Predicted labels	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$precision = \frac{TP}{TP + FP}$$

➔ How many positive predicted labels are true ?

F1-score

Confusion Matrix:

		True labels	
		Positive	Negative
Predicted labels	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$precision = \frac{TP}{TP + FP}$$

➔ How many positive predicted labels are true ?

$$recall = \frac{TP}{TP + FN}$$

➔ How many true positive labels are retrieved ?

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

F1-score

Confusion Matrix:

		True labels	
		Positive	Negative
Predicted labels	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

$$precision = \frac{TP}{TP + FP}$$

➔ How many positive predicted labels are true ?

$$recall = \frac{TP}{TP + FN}$$

➔ How many true positive labels are retrieved ?

$$F = \frac{2}{\frac{1}{recall} + \frac{1}{precision}} = \frac{2precision.recall}{recall + precision}$$

Let us apply a supervised version of RGCCA on the MDD case study

➔ See section 4 on the Rmarkdown `MDD_case_study_RGCCA`

1. Introduction of the case study

2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)

3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)

4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)

5. **Supervised analysis with RGCCA**

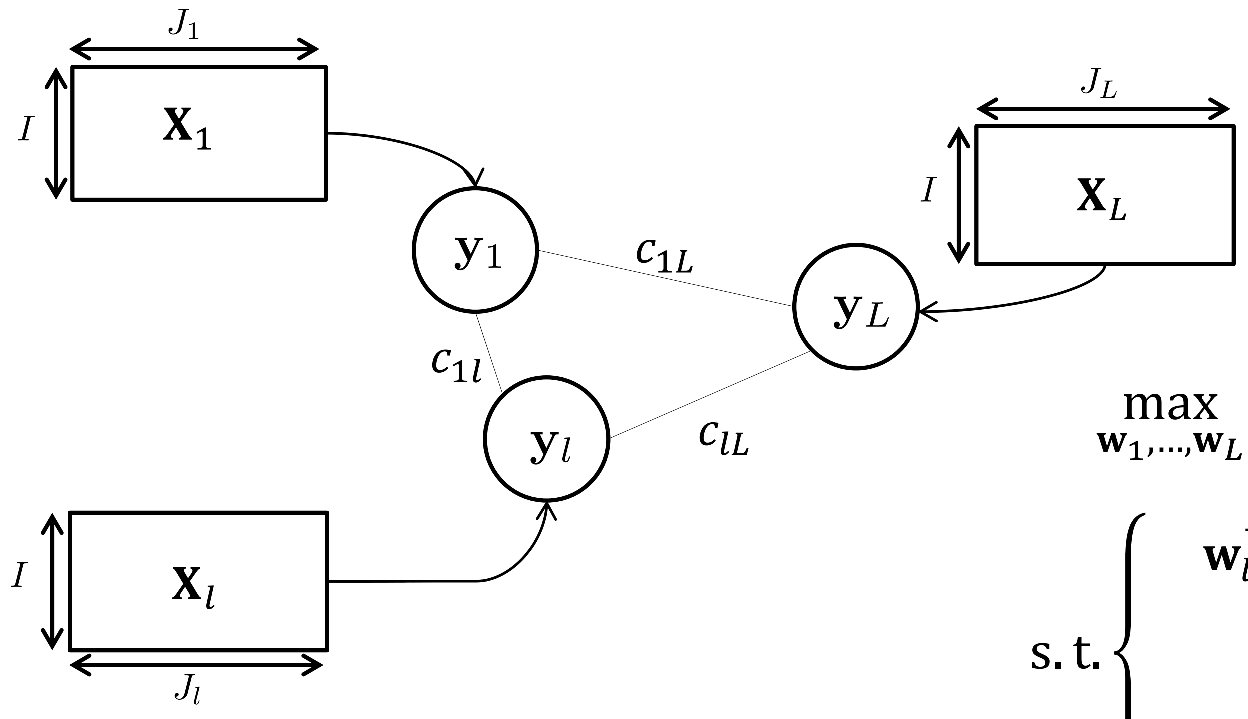
6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)

7. **The flexible Optimization Framework of RGCCA**

   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

INSTITUT FRANÇAIS DE BIOINFORMATIQUE IFB elixir FRANCE

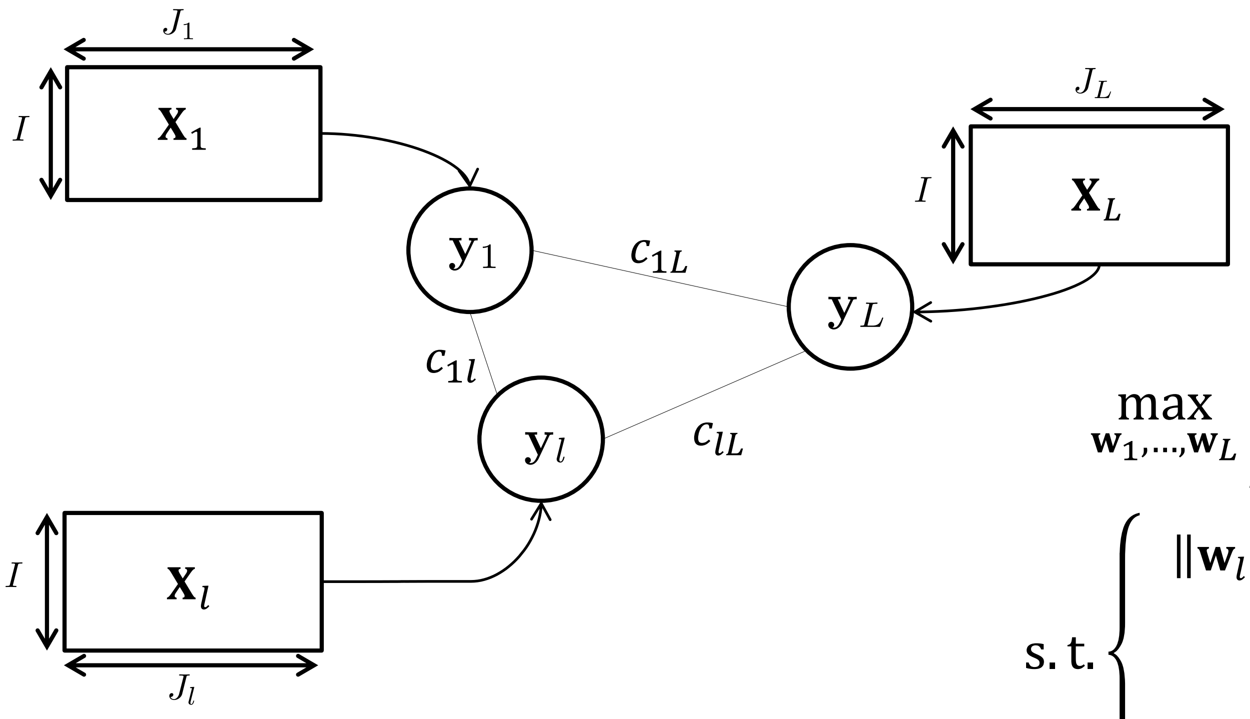
Sparse Generalized Canonical Correlation Analysis (SGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad\qquad , l = 1, \ldots, L. \\ \end{cases}$$

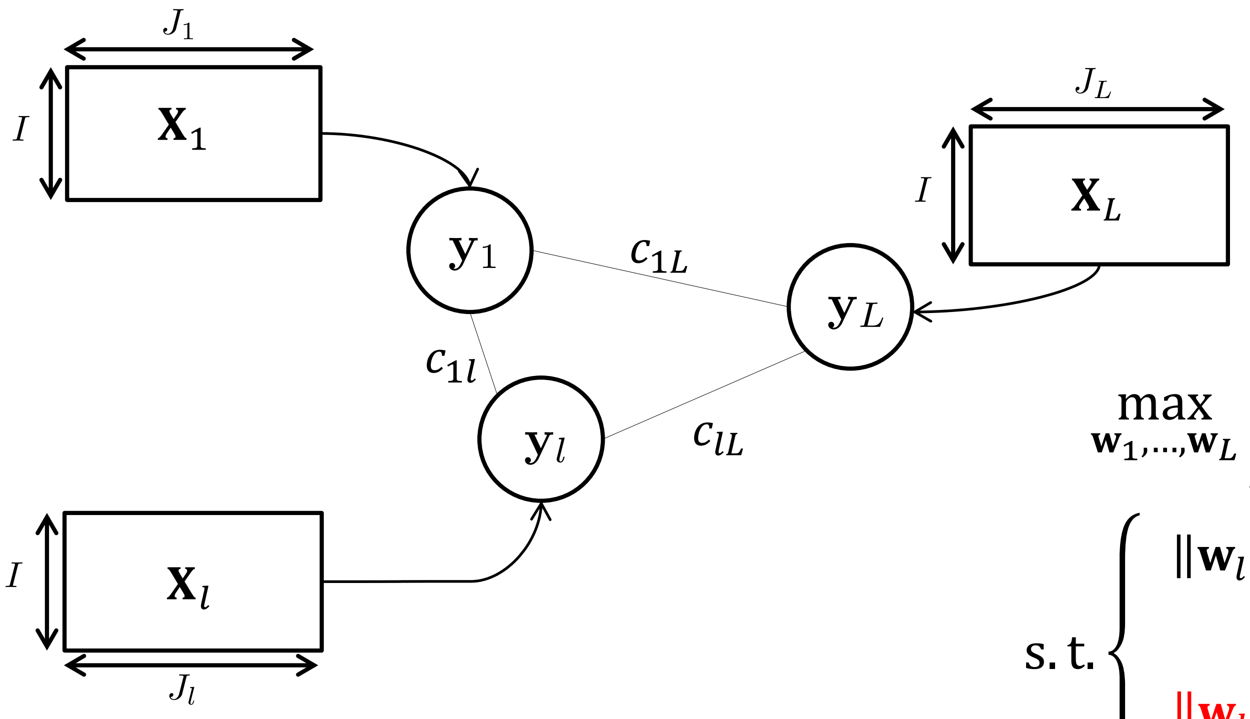
Sparse Generalized Canonical Correlation Analysis (SGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \qquad\qquad\qquad\qquad , l = 1, \ldots, L. \end{cases}$$

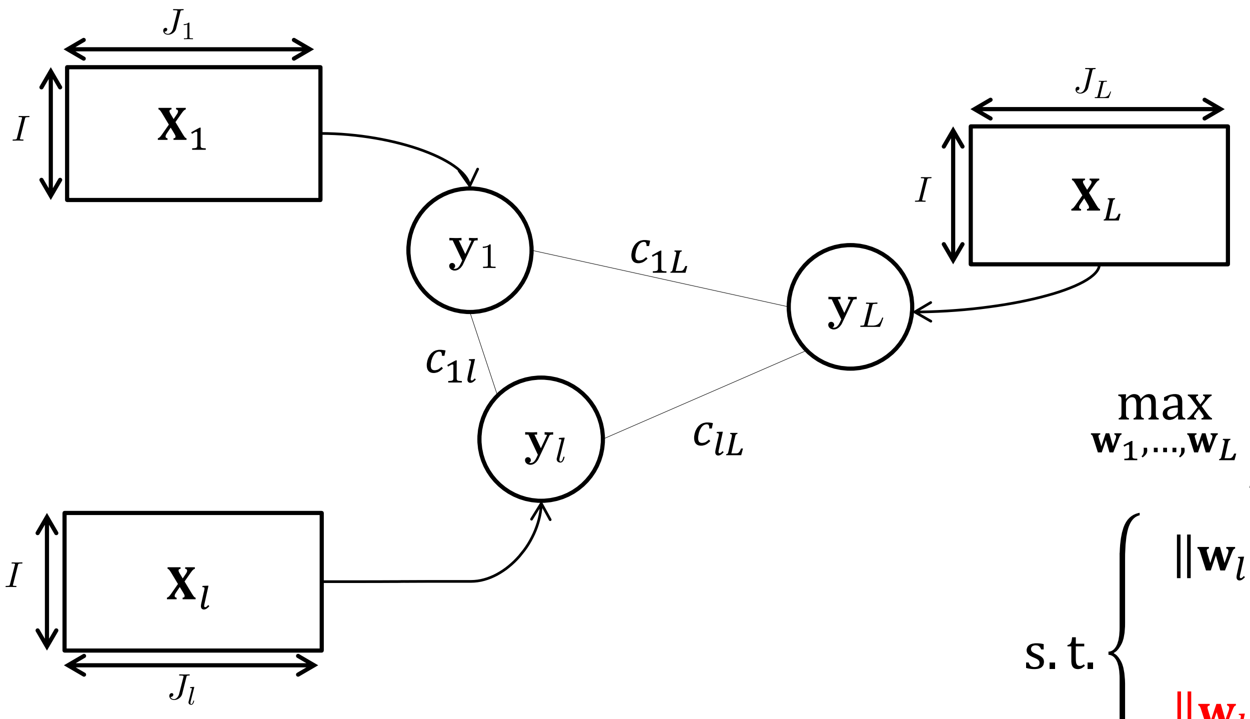
Sparse Generalized Canonical Correlation Analysis (SGCCA)



$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 = \sum_{j=1}^{J_l} |w_{lj}| \leq s_l \end{cases}, l = 1, \dots, L.$$

Sparse Generalized Canonical Correlation Analysis (SGCCA)

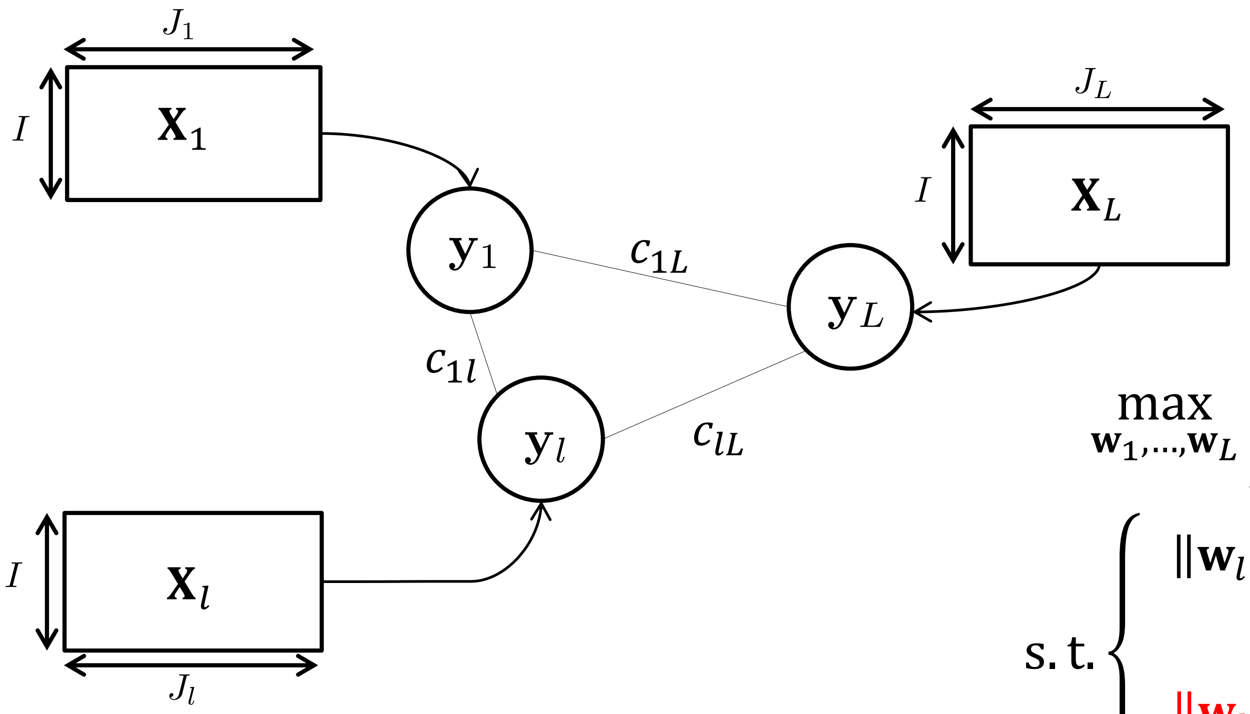


$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 = \sum_{j=1}^{J_l} |w_{lj}| \leq s_l \end{cases}, l = 1, \ldots, L.$$

The LASSO regularization allows to perform variable selection.

Sparse Generalized Canonical Correlation Analysis (SGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 = \sum_{j=1}^{J_l} |w_{lj}| \leq s_l \end{cases}, l = 1, \ldots, L.$$

Controls the level of sparsity (has to be tuned).

The LASSO regularization allows to perform variable selection.

The Variable Importance in Projection (VIP) score

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \,\mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^{2}\mathrm{AVE}\left(\mathbf{X}_{l}^{(r)}\right)\right)$$

Where:

- R is the number of extracted components.

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- R is the number of extracted components.
- $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- ❖ R is the number of extracted components.
- ❖ $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.
- ❖ RGCCA uses a deflation procedure to extract the following components.
  Thus, $\mathbf{X}_l^{(r)}$ correspond to the projection of $\mathbf{X}_l^{(r-1)}$ onto the space orthogonal to $\boldsymbol{y}_l^{(r)}$:

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- ❖ R is the number of extracted components.
- ❖ $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.
- ❖ RGCCA uses a deflation procedure to extract the following components.
  Thus, $\mathbf{X}_l^{(r)}$ correspond to the projection of $\mathbf{X}_l^{(r-1)}$ onto the space orthogonal to $\boldsymbol{y}_l^{(r)}$:

$$\mathbf{X}_l^{(r)} = \left(\mathbf{I}_{J_l} - \frac{\boldsymbol{y}_l^{(r)}{\boldsymbol{y}_l^{(r)}}^\top}{\left\|\boldsymbol{y}_l^{(r)}\right\|_2^2}\right)\mathbf{X}_l^{(r-1)}$$

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- R is the number of extracted components.
- $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.
- RGCCA uses a deflation procedure to extract the following components. Thus, $\mathbf{X}_l^{(r)}$ correspond to the projection of $\mathbf{X}_l^{(r-1)}$ onto the space orthogonal to $\boldsymbol{y}_l^{(r)}$:

$$\mathbf{X}_l^{(r)} = \left(\mathbf{I}_{J_l} - \frac{\boldsymbol{y}_l^{(r)}{\boldsymbol{y}_l^{(r)}}^\top}{\left\|\boldsymbol{y}_l^{(r)}\right\|_2^2}\right)\mathbf{X}_l^{(r-1)}$$

- Furthermore $\mathbf{X}_l^{(0)} = \mathbf{X}_l$

The Variable Importance in Projection (VIP) score

$$\text{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \text{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- R is the number of extracted components.
- $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.
- RGCCA uses a deflation procedure to extract the following components. Thus, $\mathbf{X}_l^{(r)}$ correspond to the projection of $\mathbf{X}_l^{(r-1)}$ onto the space orthogonal to $\boldsymbol{y}_l^{(r)}$:

$$\mathbf{X}_l^{(r)} = \left(\mathbf{I}_{J_l} - \frac{\boldsymbol{y}_l^{(r)}{\boldsymbol{y}_l^{(r)}}^\top}{\left\|\boldsymbol{y}_l^{(r)}\right\|_2^2}\right)\mathbf{X}_l^{(r-1)}$$

- Furthermore $\mathbf{X}_l^{(0)} = \mathbf{X}_l$
- The Average Variance Explained (AVE) associated with $\boldsymbol{y}_l^{(r)}$ is:

The Variable Importance in Projection (VIP) score

$$\mathrm{VIP}\left(\mathbf{x}_{lj}\right) = \frac{1}{R}\sum_{r=1}^{R}\left({w_{lj}^{(r)}}^2 \mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right)\right)$$

Where:

- ❖ R is the number of extracted components.
- ❖ $\mathbf{X}_l = \left[\mathbf{x}_{l1}, \dots, \mathbf{x}_{lJ_l}\right]$ and $\mathbf{w}_l = \left[w_{l1}, \dots, w_{lJ_l}\right]^\top$.
- ❖ RGCCA uses a deflation procedure to extract the following components.
  Thus, $\mathbf{X}_l^{(r)}$ correspond to the projection of $\mathbf{X}_l^{(r-1)}$ onto the space orthogonal to $\boldsymbol{y}_l^{(r)}$:

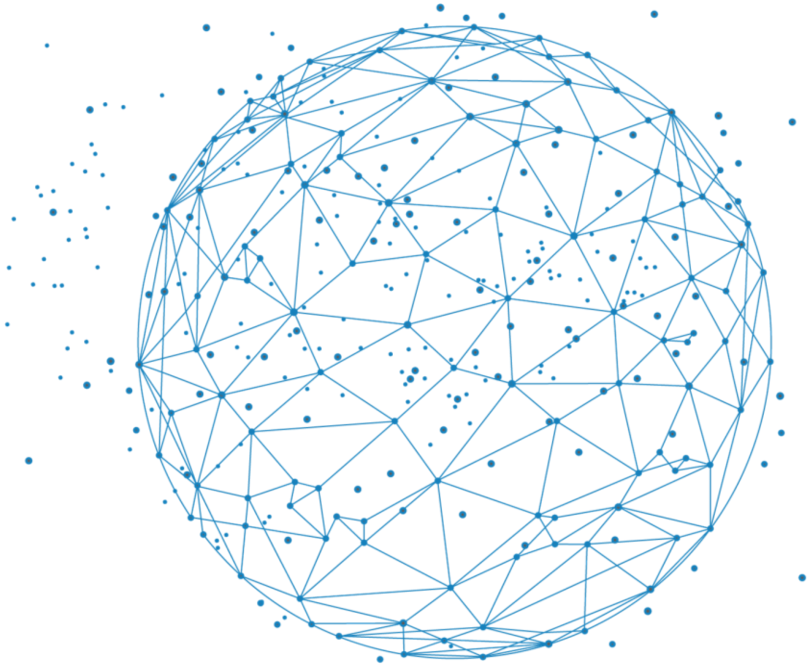
$$\mathbf{X}_l^{(r)} = \left(\mathbf{I}_{J_l} - \frac{\boldsymbol{y}_l^{(r)}{\boldsymbol{y}_l^{(r)}}^\top}{\left\|\boldsymbol{y}_l^{(r)}\right\|_2^2}\right)\mathbf{X}_l^{(r-1)}$$

- ❖ Furthermore $\mathbf{X}_l^{(0)} = \mathbf{X}_l$
- ❖ The Average Variance Explained (AVE) associated with $\boldsymbol{y}_l^{(r)}$ is:

$$\mathrm{AVE}\left(\mathbf{X}_l^{(r)}\right) = \frac{1}{\left\|\mathbf{X}_l^{(r)}\right\|_F^2}\sum_{j=1}^{J_l}\left(\mathrm{var}\left(\mathbf{x}_{lj}^{(r)}\right) \times \mathrm{cor}^2\left(\mathbf{x}_{lj}^{(r)}, \boldsymbol{y}_l^{(r+1)}\right)\right)$$

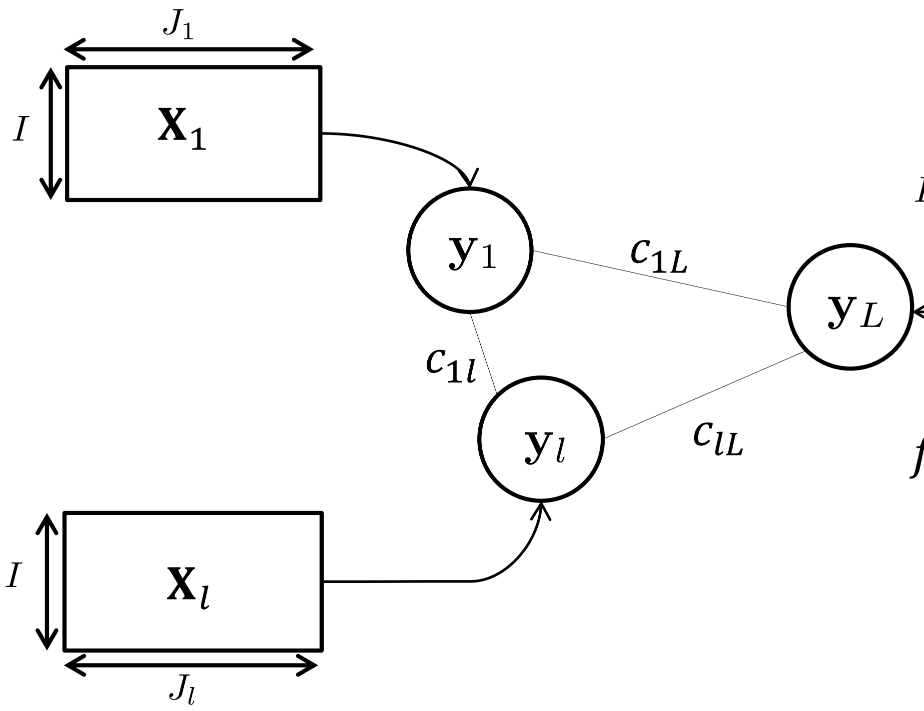
Let us apply both an unsupervised/supervised version of SGCCA on the MDD case study

➔ See section 5 & 6 on the Rmarkdown `MDD_case_study_RGCCA`



1. Introduction of the case study
2. Unsupervised analysis with one-block: Principal Component Analysis (PCA)
3. Unsupervised analysis with two-blocks:
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)
4. Unsupervised analysis with L-blocks:
   Regularized Generalized Canonical Correlation Analysis (RGCCA)
5. Supervised analysis with RGCCA
6. Variable selection in RGCCA:
   Sparse Generalized Canonical Correlation Analysis (SGCCA)
7. **The flexible Optimization Framework of RGCCA**
   - **The general principal**
   - Extension to multi-way analysis
   - From Sequential to Global

Principle of the RGCCA Optimization Algorithm

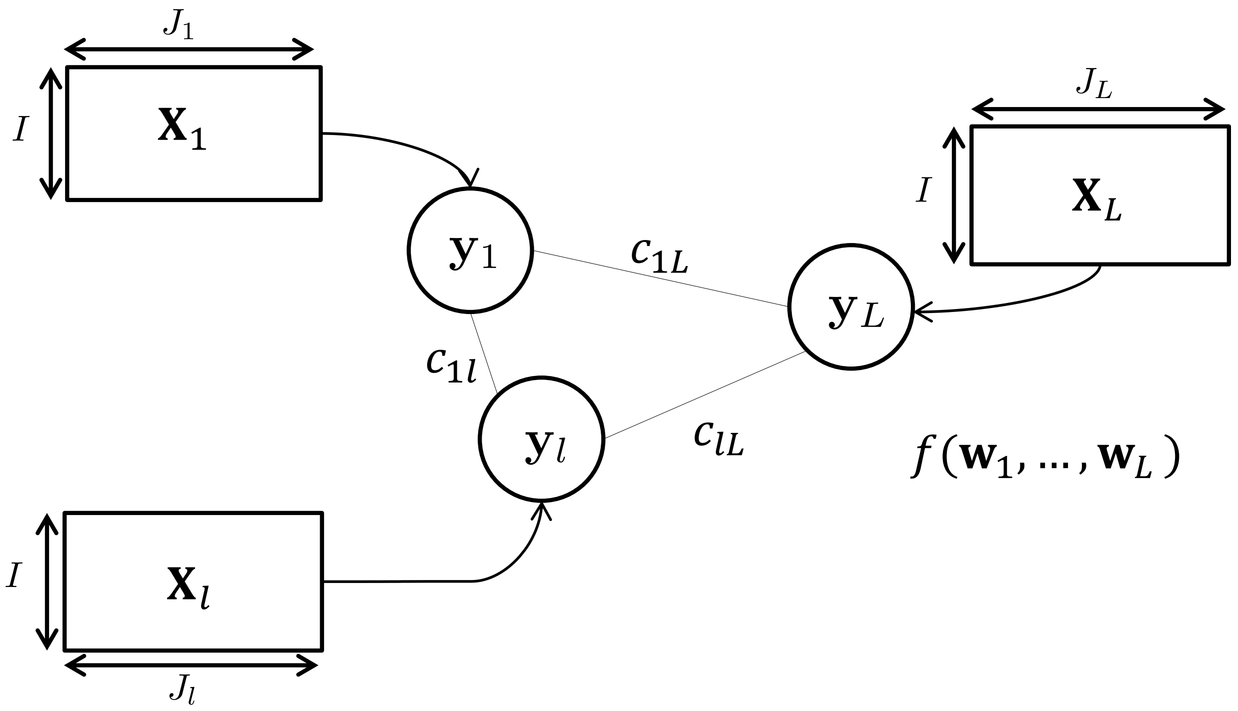


$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases} , l = 1, \dots, L.$$

Principle of the RGCCA Optimization Algorithm

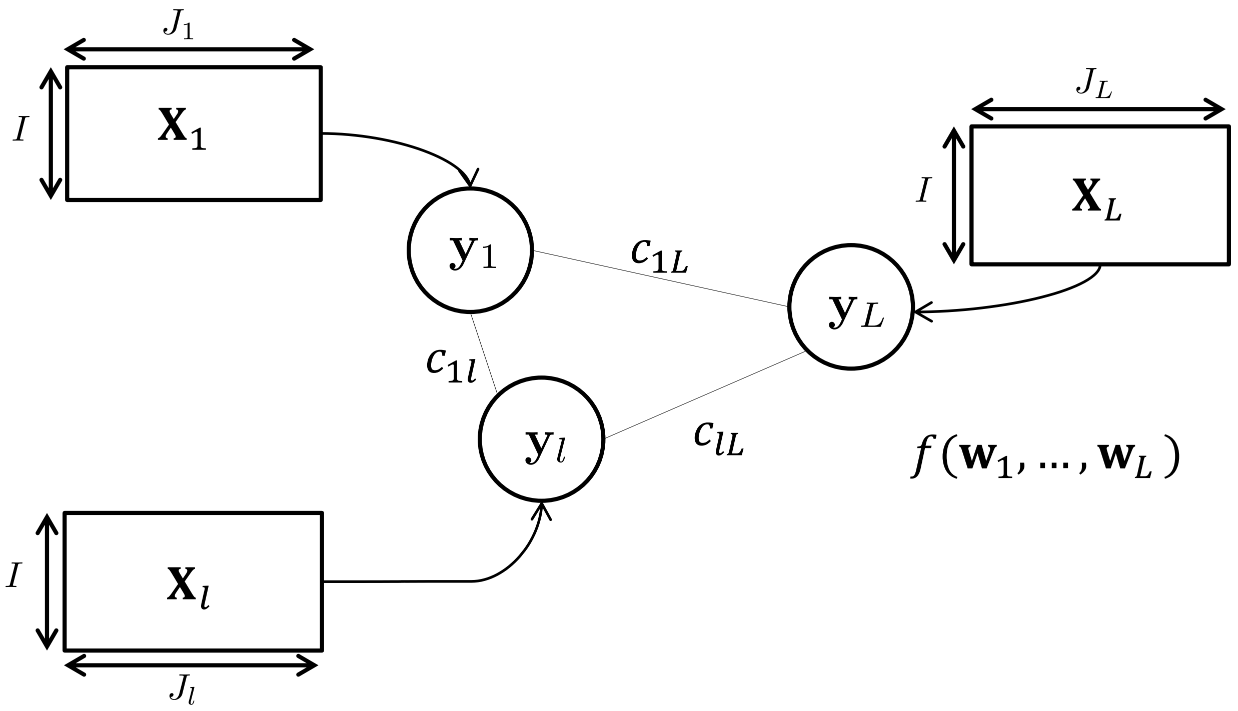


$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases}, l = 1, \dots, L.$$

Principle of the RGCCA Optimization Algorithm



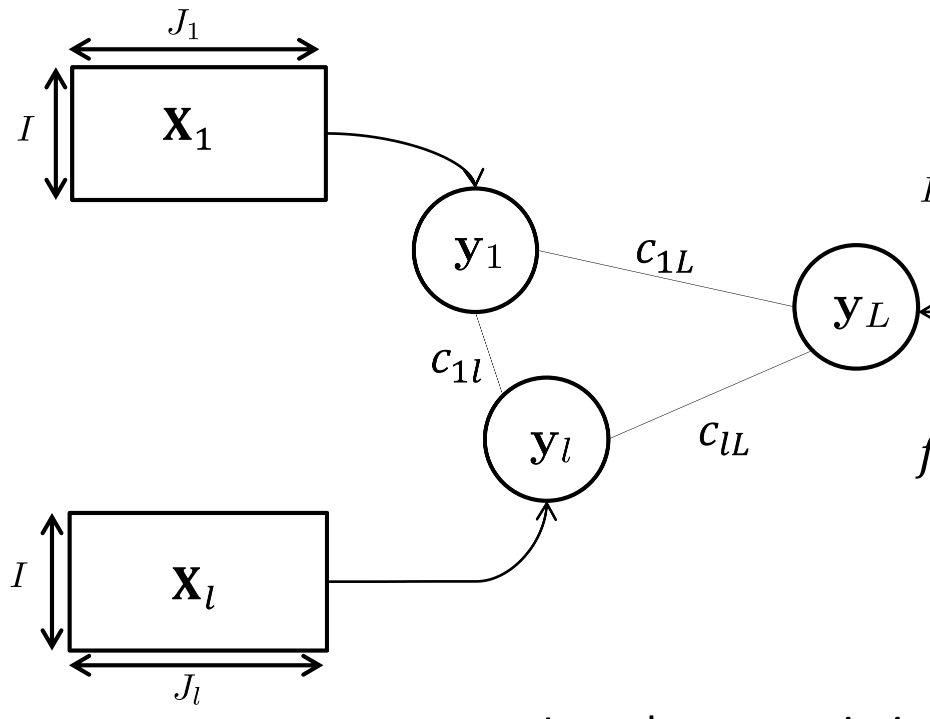
$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad , l = 1, \dots, L. \end{cases}$$

Principle of the RGCCA Optimization Algorithm



$f(\mathbf{w}_1, \ldots, \mathbf{w}_L)$

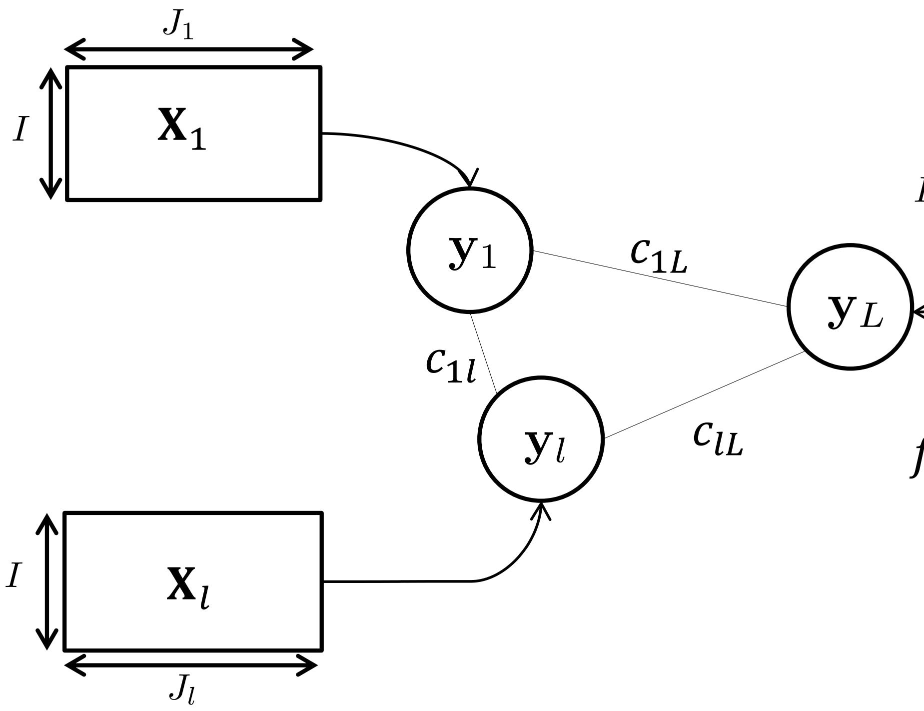
$f(\mathbf{w}_1, \ldots, \mathbf{w}_L)$

$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases}, l = 1, \ldots, L.$$

In order to maximize the multi-convex function $f(\mathbf{w}_1, \ldots, \mathbf{w}_L)$,
two key ingredients are used:

Principle of the RGCCA Optimization Algorithm



$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

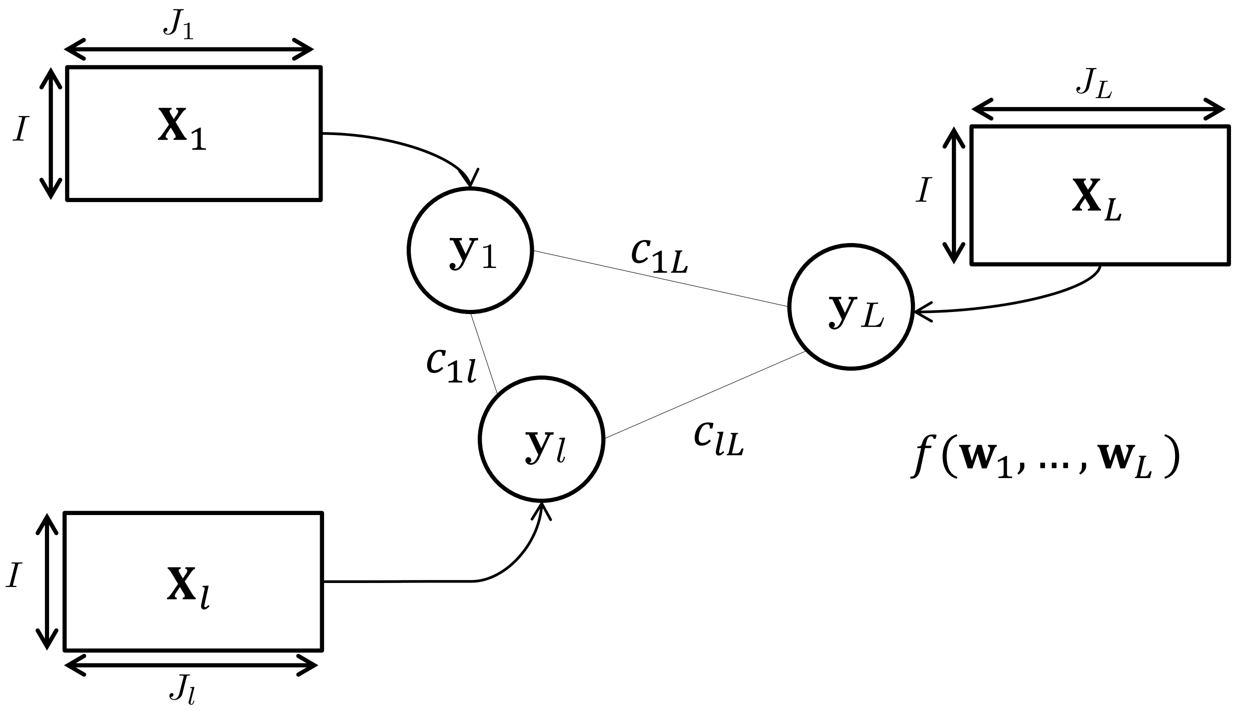
$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases}, l = 1, \dots, L.$$

In order to maximize the multi-convex function $f(\mathbf{w}_1, \dots, \mathbf{w}_L)$, two key ingredients are used:

Block Coordinate Ascent (BCA).

Principle of the RGCCA Optimization Algorithm



$f(\mathbf{w}_1, \dots, \mathbf{w}_L)$

$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases}, l = 1, \dots, L.$$

In order to maximize the multi-convex function $f(\mathbf{w}_1, \dots, \mathbf{w}_L)$, two key ingredients are used:

- Block Coordinate Ascent (BCA).
- Minorize-Maximize (MM) principle.

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \ldots, \mathbf{w}_L^s)$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} \; f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} \; f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \ldots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \ldots, \mathbf{w}_L^s)$$

$$\mathbf{w}_1^{s+1}$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s)$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\text{argmax}} \; f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$\rightarrow \mathbf{w}_1^{s+1}$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\text{argmax}} \; f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s)$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_2^{s+1}$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_2^{s+1}$$

$$\vdots$$

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_l^{s+1}$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_2^{s+1}$$

$$\vdots$$

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_l^{s+1}$$

$$\vdots$$

$$\underset{\mathbf{w}_L, \mathbf{w}_L^\top \mathbf{M}_L \mathbf{w}_L = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{L-1}^{s+1}, \mathbf{w}_L^s) \rightarrow \mathbf{w}_L^{s+1}$$

Block Coordinate Ascent

$$\mathbf{w}^s = (\mathbf{w}_1^s, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s)$$

$$\underset{\mathbf{w}_1, \mathbf{w}_1^\top \mathbf{M}_1 \mathbf{w}_1 = 1}{\operatorname{argmax}} f(\mathbf{w}_1, \mathbf{w}_2^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_1^{s+1}$$

$$\underset{\mathbf{w}_2, \mathbf{w}_2^\top \mathbf{M}_2 \mathbf{w}_2 = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \mathbf{w}_2, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_2^{s+1}$$

$$\vdots$$

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^s, \dots, \mathbf{w}_L^s) \rightarrow \mathbf{w}_l^{s+1}$$

$$\vdots$$

$$\underset{\mathbf{w}_L, \mathbf{w}_L^\top \mathbf{M}_L \mathbf{w}_L = 1}{\operatorname{argmax}} f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{L-1}^{s+1}, \mathbf{w}_L^s) \rightarrow \mathbf{w}_L^{s+1}$$

$$\mathbf{w}^{s+1} = (\mathbf{w}_1^{s+1}, \mathbf{w}_2^{s+1}, \dots, \mathbf{w}_L^{s+1})$$

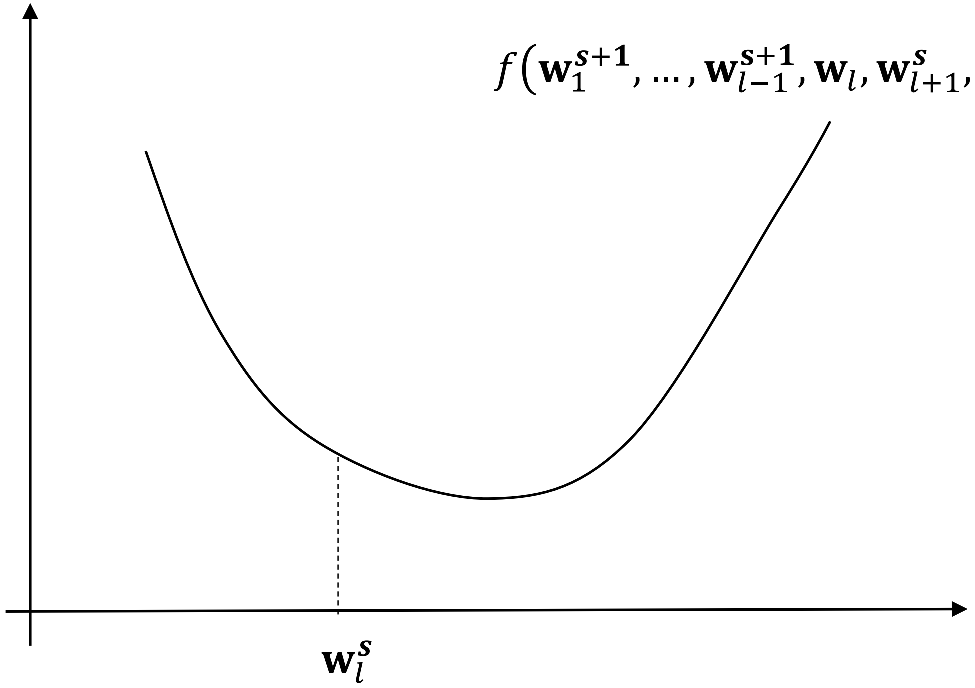
INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \quad f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Minorize-Maximize (MM) principle

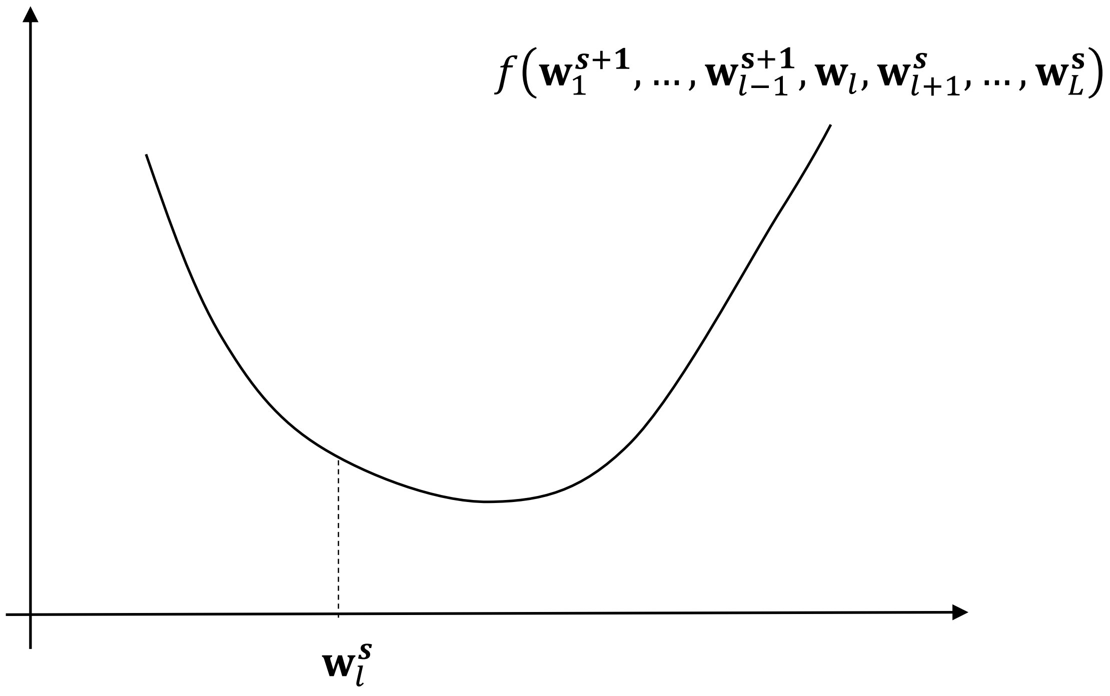
$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$



Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \quad f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

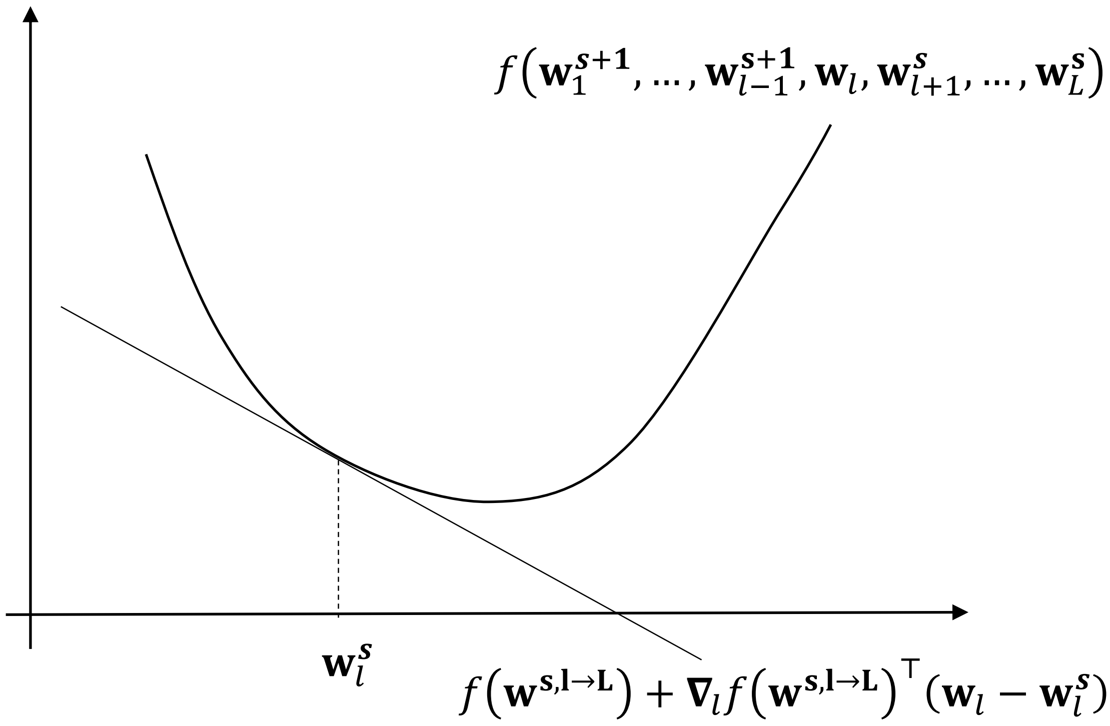
Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\rightarrow\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{\mathbf{s+1}}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{\mathbf{s}}\right)$



Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \quad f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s,l\to L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

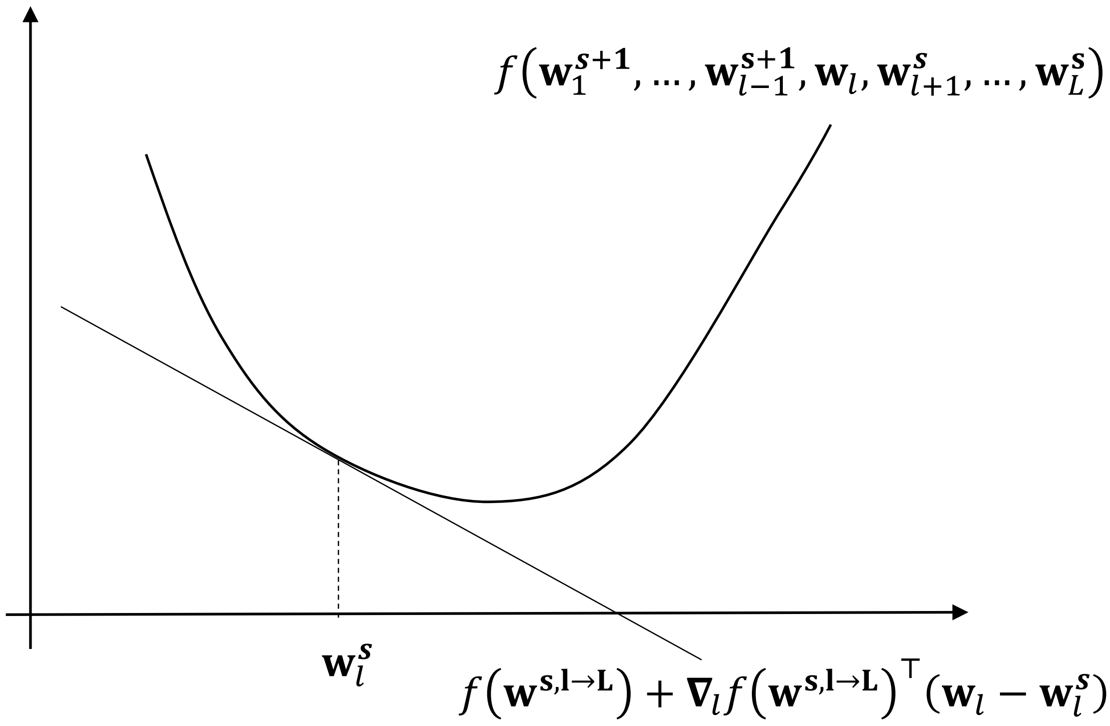


Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\rightarrow\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\rightarrow\mathbf{L}}\right)^\top \mathbf{w}_l$$

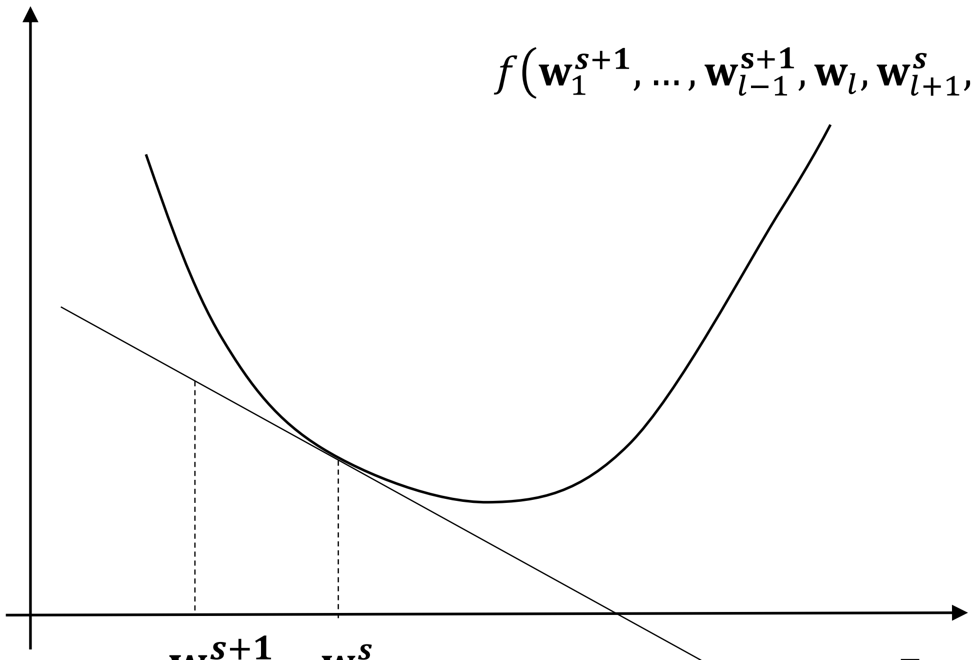


Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \quad f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{\mathbf{s+1}}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \quad \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \mathbf{w}_l$$



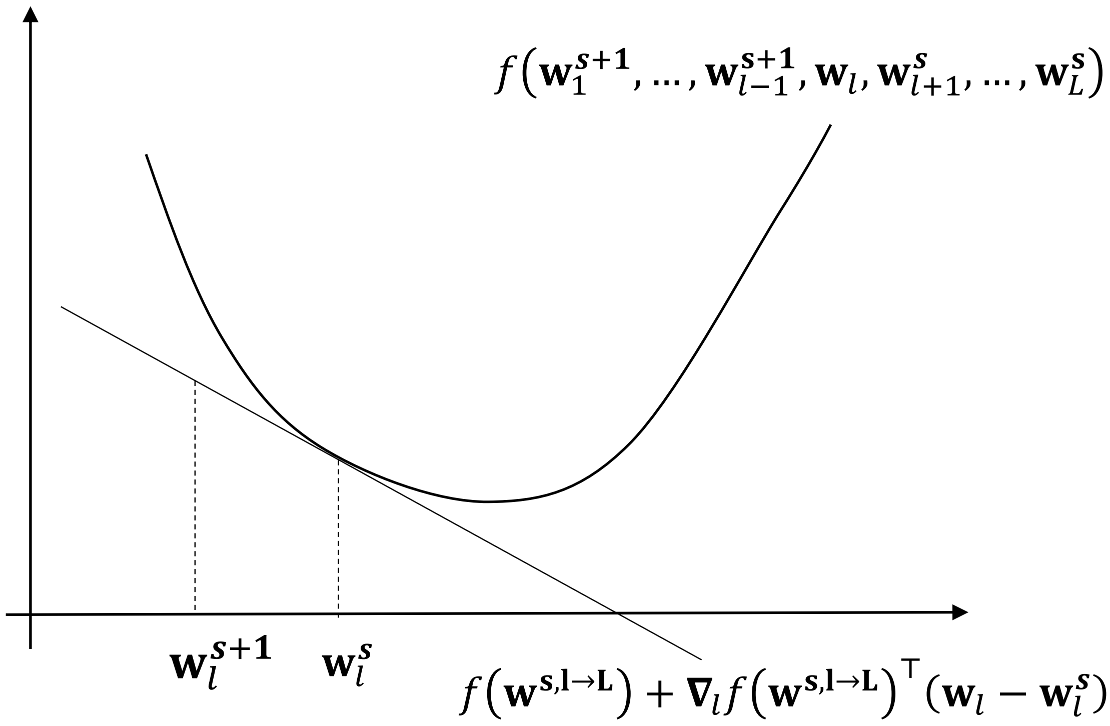
Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l,\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l=1}{\text{argmax}} \quad f(\mathbf{w}_1^{s+1},\dots,\mathbf{w}_{l-1}^{s+1},\mathbf{w}_l,\mathbf{w}_{l+1}^{s},\dots,\mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s,l\to L}} = \left(\mathbf{w}_1^{s+1},\dots,\mathbf{w}_{l-1}^{\mathbf{s+1}},\mathbf{w}_l^{s},\mathbf{w}_{l+1}^{s},\dots,\mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l,\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l=1}{\text{argmax}} \quad \nabla_l f\left(\mathbf{w}^{\mathbf{s,l\to L}}\right)^\top \mathbf{w}_l$$

$$f\left(\mathbf{w}^{\mathbf{s,l\to L}}\right) = f\left(\mathbf{w}^{\mathbf{s,l\to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l\to L}}\right)^\top (\mathbf{w}_l^s - \mathbf{w}_l^s)$$



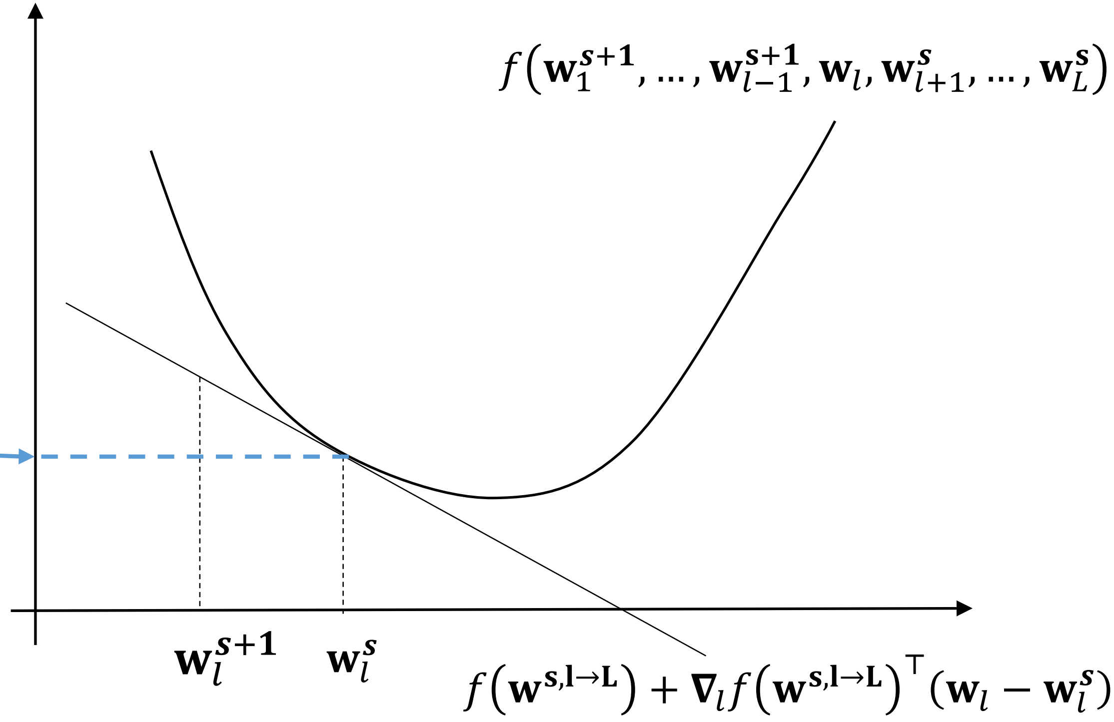
Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{\mathbf{s+1}}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \mathbf{w}_l$$

$$f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) = f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top (\mathbf{w}_l^s - \mathbf{w}_l^s)$$



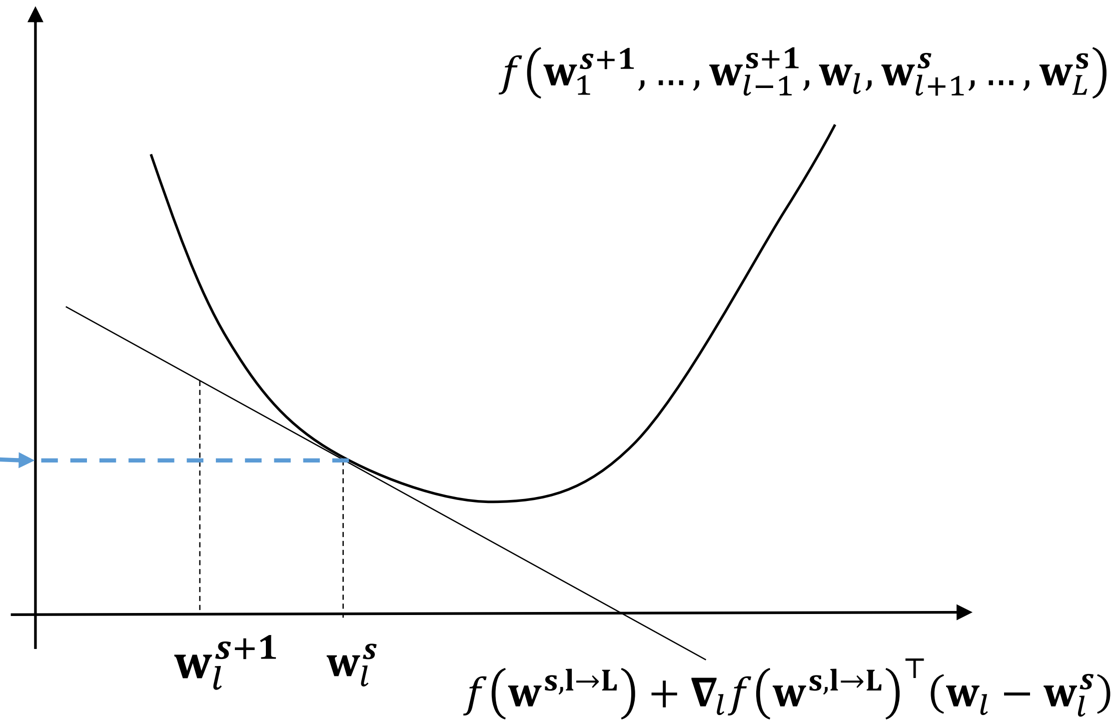
Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l,\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l=1}{\operatorname{argmax}} \; f\left(\mathbf{w}_1^{s+1},\dots,\mathbf{w}_{l-1}^{s+1},\mathbf{w}_l,\mathbf{w}_{l+1}^{s},\dots,\mathbf{w}_L^{s}\right)$$

Let us introduce: $\mathbf{w}^{\mathrm{s,l\to L}} = \left(\mathbf{w}_1^{s+1},\dots,\mathbf{w}_{l-1}^{s+1},\mathbf{w}_l^{s},\mathbf{w}_{l+1}^{s},\dots,\mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l,\mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l=1}{\operatorname{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned} f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) &= f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right) \\ &\le f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right) \end{aligned}$$



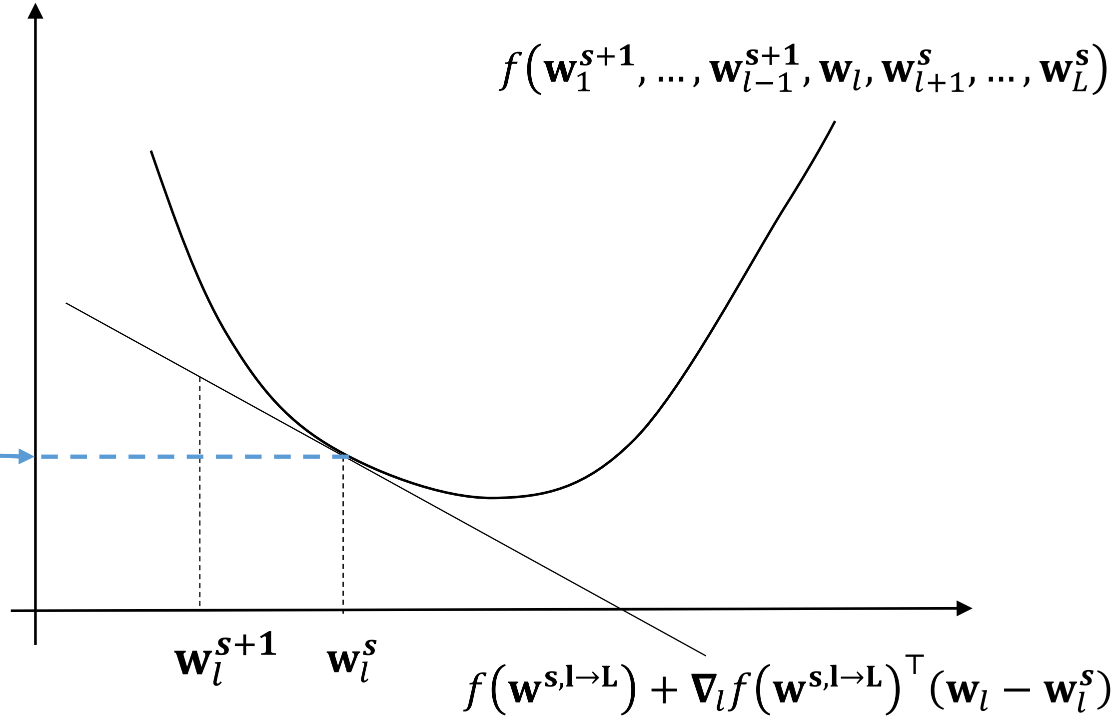
Minorize-Maximize (MM) principle

$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned} f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) &= f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top (\mathbf{w}_l^s - \mathbf{w}_l^s) \\ &\leq f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top (\mathbf{w}_l^{s+1} - \mathbf{w}_l^s) \end{aligned}$$



Minorize-Maximize (MM) principle

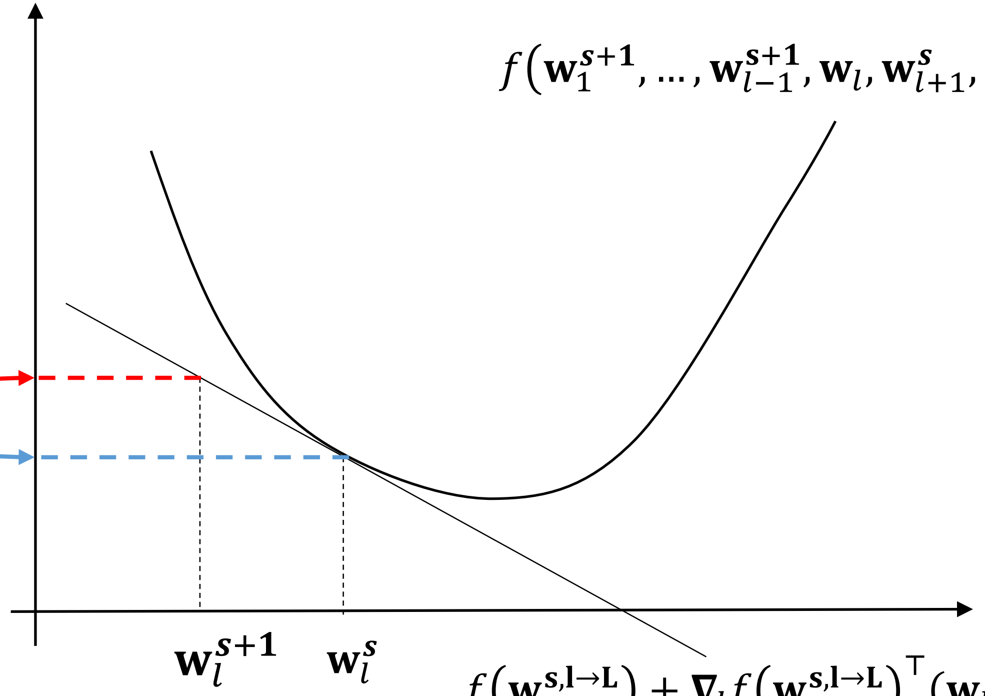
$$\underset{\mathbf{w}_l,\, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$$

Let us introduce: $\mathbf{w}^{\mathbf{s,l \to L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l,\, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \mathbf{w}_l$$

$$f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) = f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right)$$

$$\leq f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right)$$

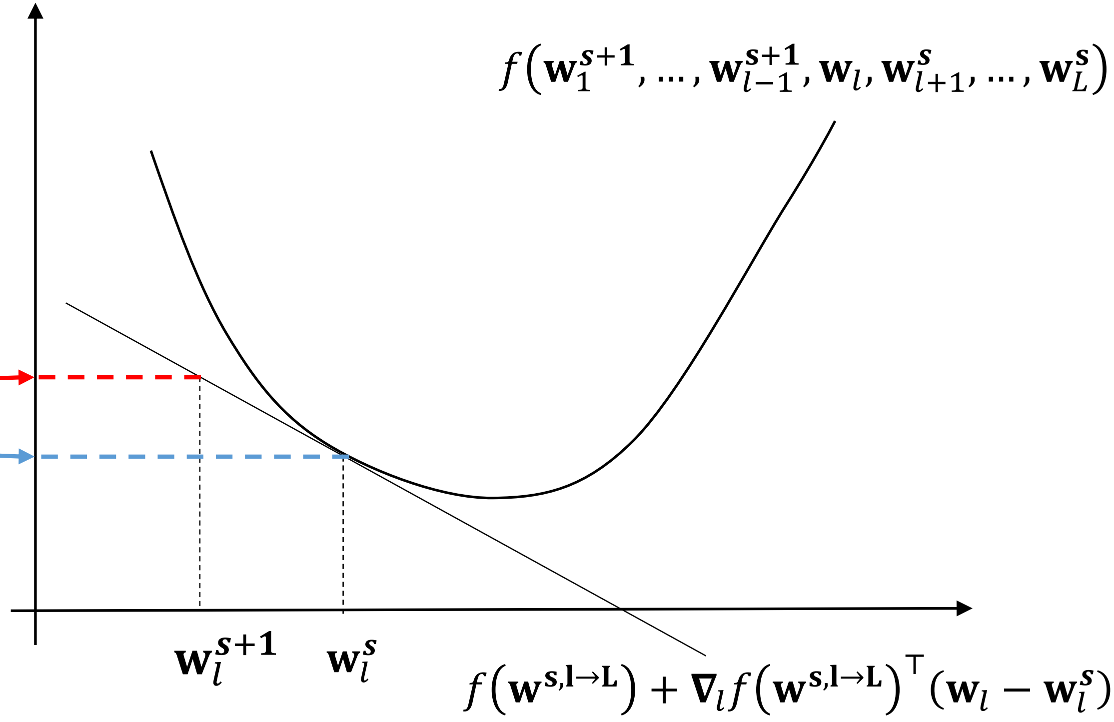


$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$$

Let us introduce: $\mathbf{w}^{\mathrm{s,l\to L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\operatorname{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned} f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) &= f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathrm{s,l\to L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s+1}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right) \end{aligned}$$



Minorize-Maximize (MM) principle

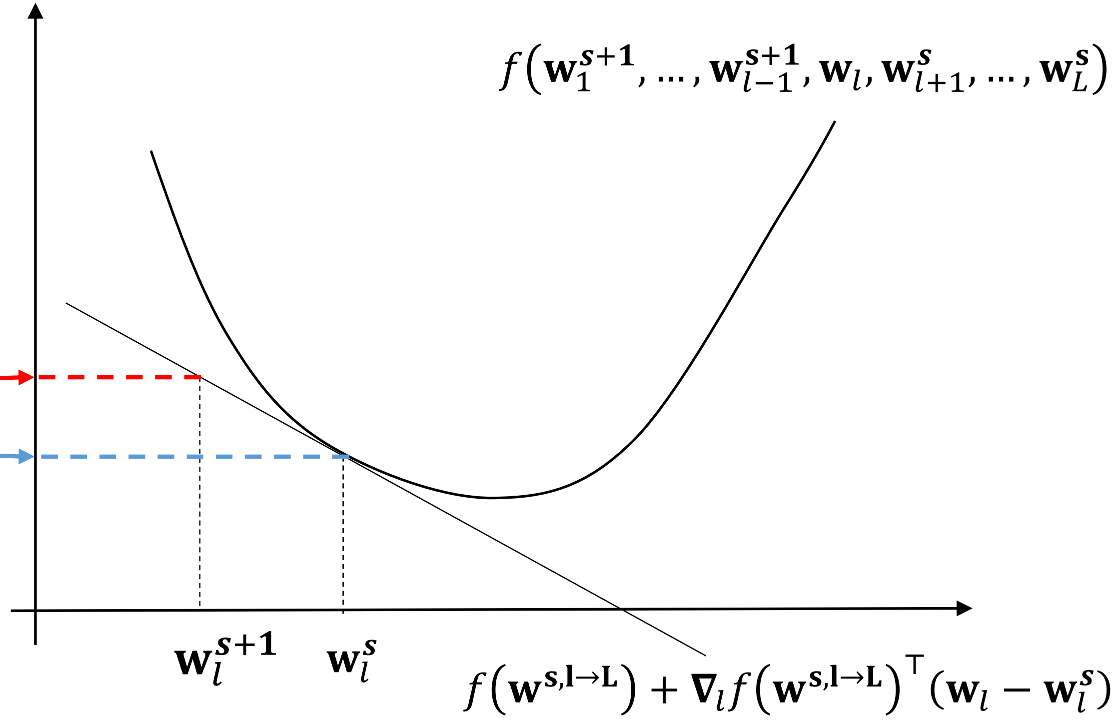
$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; f\left(\mathbf{w}_1^{s+1}, \ldots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \ldots, \mathbf{w}_L^{s}\right)$$

Let us introduce: $\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}} = \left(\mathbf{w}_1^{s+1}, \ldots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \ldots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned} f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) &= f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s},\mathbf{l}\to\mathbf{L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}_1^{s+1}, \ldots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s+1}, \mathbf{w}_{l+1}^{s}, \ldots, \mathbf{w}_L^{s}\right) \end{aligned}$$

Comes from the multi-convexity of f and so the convexity of $f\left(\mathbf{w}_1^{s+1}, \ldots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \ldots, \mathbf{w}_L^{s}\right)$



IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE elixir FRANCE

Minorize-Maximize (MM) principle

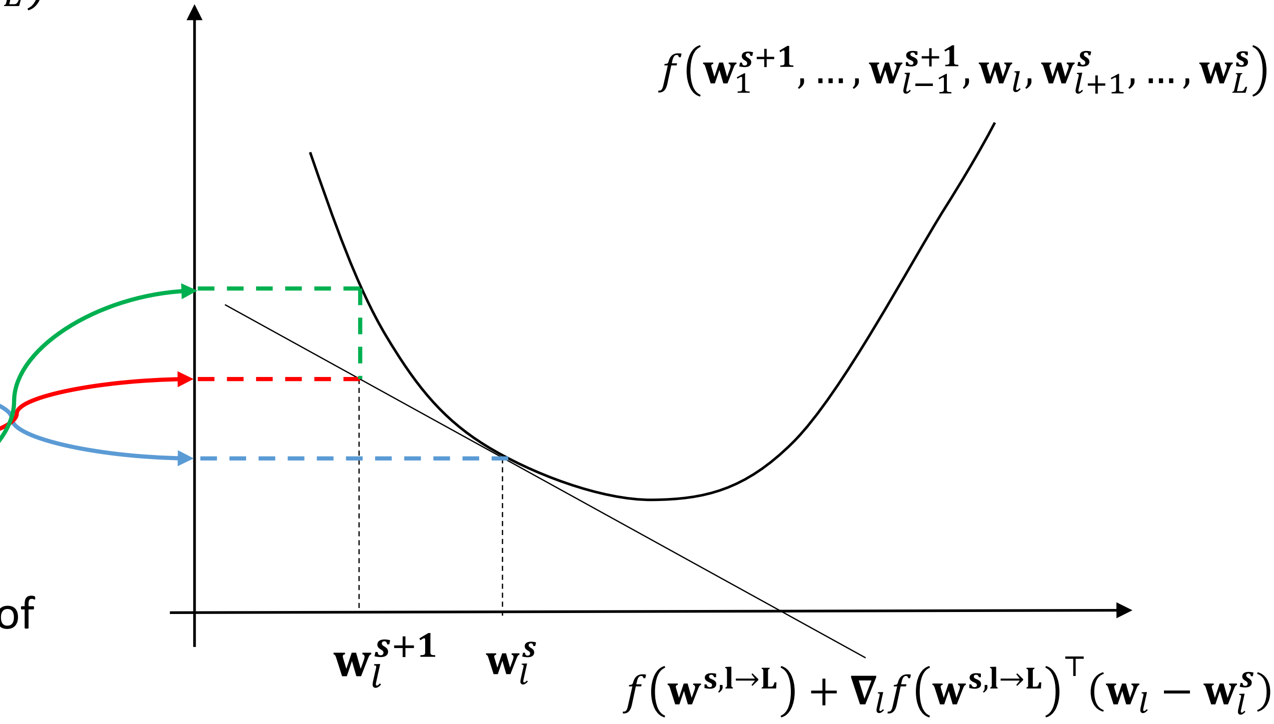
$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s,l \to L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{\mathbf{s+1}}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned}
f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) &= f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right) \\
&\leq f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right) \\
&\leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s+1}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)
\end{aligned}$$

Comes from the multi-convexity of f and so the convexity of $f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$



Minorize-Maximize (MM) principle

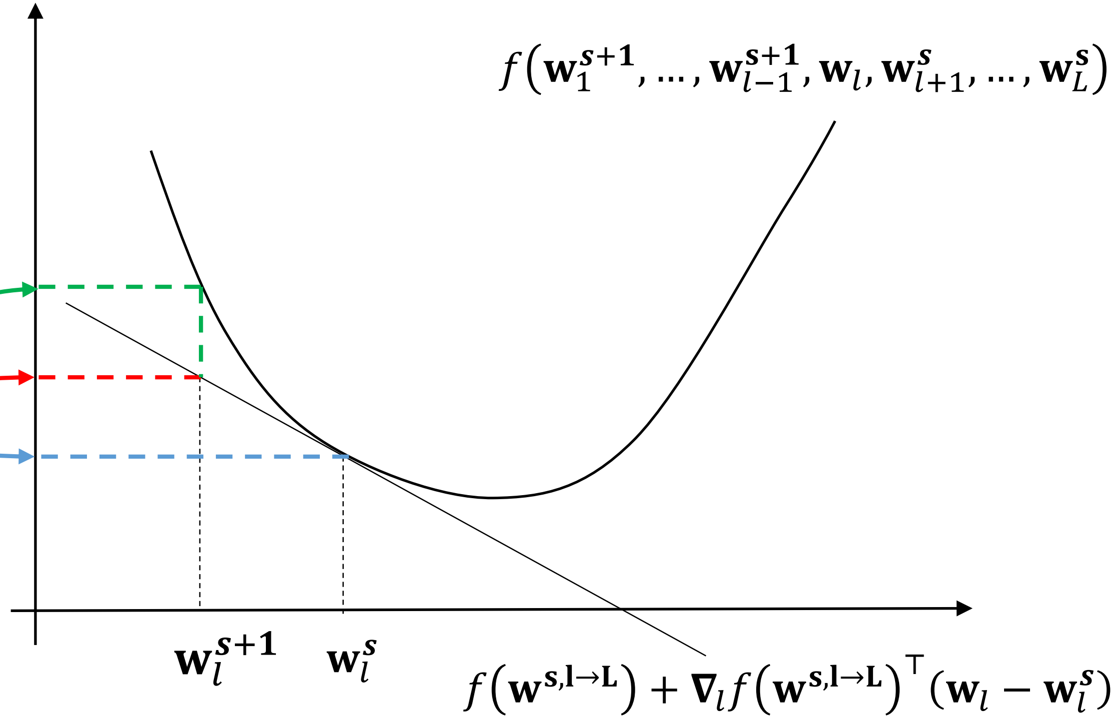
$$\underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$$

Let us introduce: $\mathbf{w}^{\mathbf{s,l \to L}} = \left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

$$\mathbf{w}_l^{s+1} = \underset{\mathbf{w}_l, \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1}{\text{argmax}} \; \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \mathbf{w}_l$$

$$\begin{aligned} f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) &= f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right) + \nabla_l f\left(\mathbf{w}^{\mathbf{s,l \to L}}\right)^\top \left(\mathbf{w}_l^{s+1} - \mathbf{w}_l^{s}\right) \\ &\leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s+1}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right) \end{aligned}$$

Comes from the multi-convexity of f and so the convexity of $f(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s})$



In the end $f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right) \leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_{l-1}^{s+1}, \mathbf{w}_l^{s+1}, \mathbf{w}_{l+1}^{s}, \dots, \mathbf{w}_L^{s}\right)$

IFB INSTITUT FRANÇAIS DE BIOINFORMATIQUE · elixir FRANCE

Properties of this Master Algorithm

Properties of this Master Algorithm

- Monotone convergence of the algorithm:

$$f(\mathbf{w}_1^s, \dots, \mathbf{w}_L^s) \leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_L^{s+1}\right)$$

Properties of this Master Algorithm

❖ Monotone convergence of the algorithm:

$$f(\mathbf{w}_1^s, \dots, \mathbf{w}_L^s) \leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_L^{s+1}\right)$$

In addition, assuming uniqueness of the solution of the MM step, the following properties hold:

Properties of this Master Algorithm

- Monotone convergence of the algorithm:

$$f(\mathbf{w}_1^s, \dots, \mathbf{w}_L^s) \leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_L^{s+1}\right)$$

In addition, assuming uniqueness of the solution of the MM step, the following properties hold:

- The sequence $\{\mathbf{w}^s\}$is asymptotically regular:

$$\lim_{s \to +\infty} \|\mathbf{w}^{s+1} - \mathbf{w}^s\|_2 = 0$$

Properties of this Master Algorithm

- Monotone convergence of the algorithm:
  $$f(\mathbf{w}_1^{s}, \dots, \mathbf{w}_L^{s}) \leq f\left(\mathbf{w}_1^{s+1}, \dots, \mathbf{w}_L^{s+1}\right)$$

In addition, assuming uniqueness of the solution of the MM step, the following properties hold:

- The sequence $\{\mathbf{w}^{s}\}$is asymptotically regular:
  $$\lim_{s \to +\infty} \|\mathbf{w}^{s+1} - \mathbf{w}^{s}\|_2 = 0$$

- At convergence, a stationnary point is obtained.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

RGCCA framework - State of the Art of the **package**

Core Optimization Problem

Constraints

Constraints

RGCCA framework - State of the Art of the **package**

Core Optimization Problem	
$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$	

Constraints

Constraints

RGCCA framework - State of the Art of the **package**

		Core Optimization Problem				
		$\max \nabla_l f(\mathbf{w}^s)^\top \mathbf{w}_l$				
Constraints	$\mathbf{w}_l \in \omega_l$					Constraints

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

		Core Optimization Problem			
		$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017)

	Core Optimization Problem		
	$\max \nabla_l f(\mathbf{w}^s)^\top \mathbf{w}_l$		
$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]		
$\begin{cases} \lVert\mathbf{w}_l\rVert_2^2 = 1 \\ \lVert\mathbf{w}_l\rVert_1 \leq s_l \end{cases}$			

Constraints

Constraints

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017)

RGCCA framework - State of the Art of the **package**

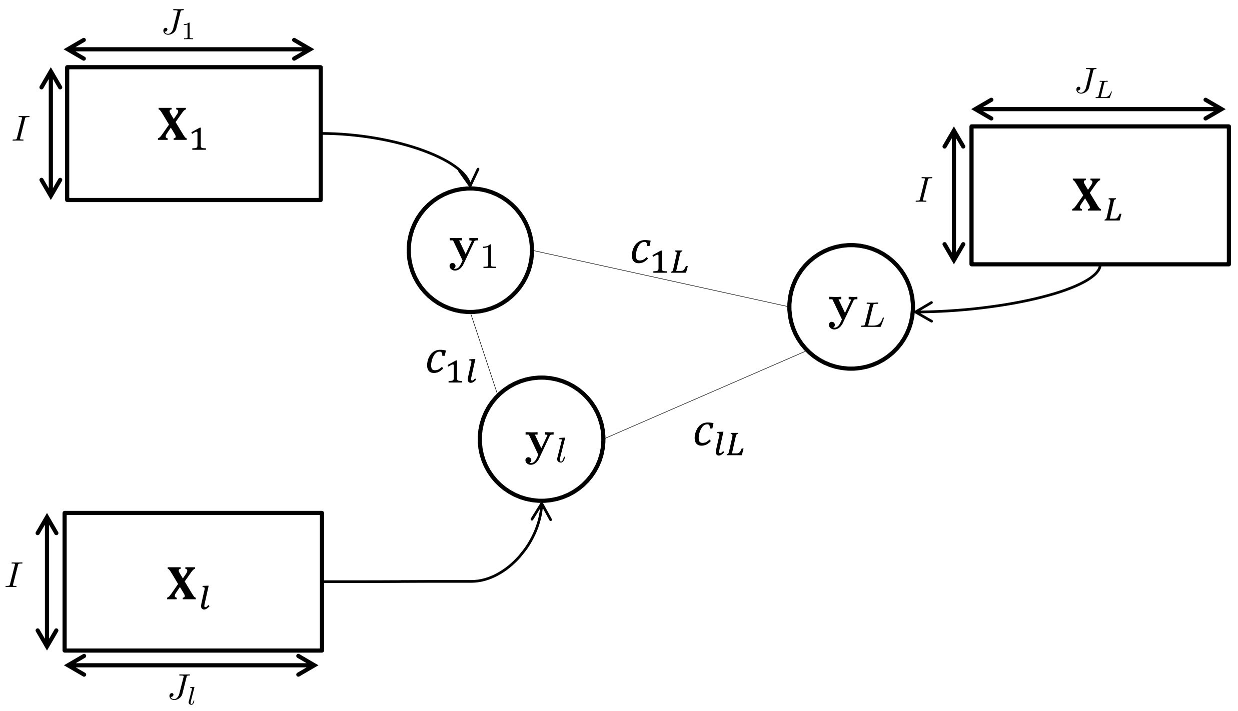
		Core Optimization Problem			
		$\max \nabla_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \lVert\mathbf{w}_l\rVert_2^2 = 1 \\ \lVert\mathbf{w}_l\rVert_1 \leq s_l \end{cases}$	SGCCA[3]			

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014)

1. Introduction of the case study
2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)
3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)
4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)
5. Supervised analysis with RGCCA
6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)
7. **The flexible Optimization Framework of RGCCA**
   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

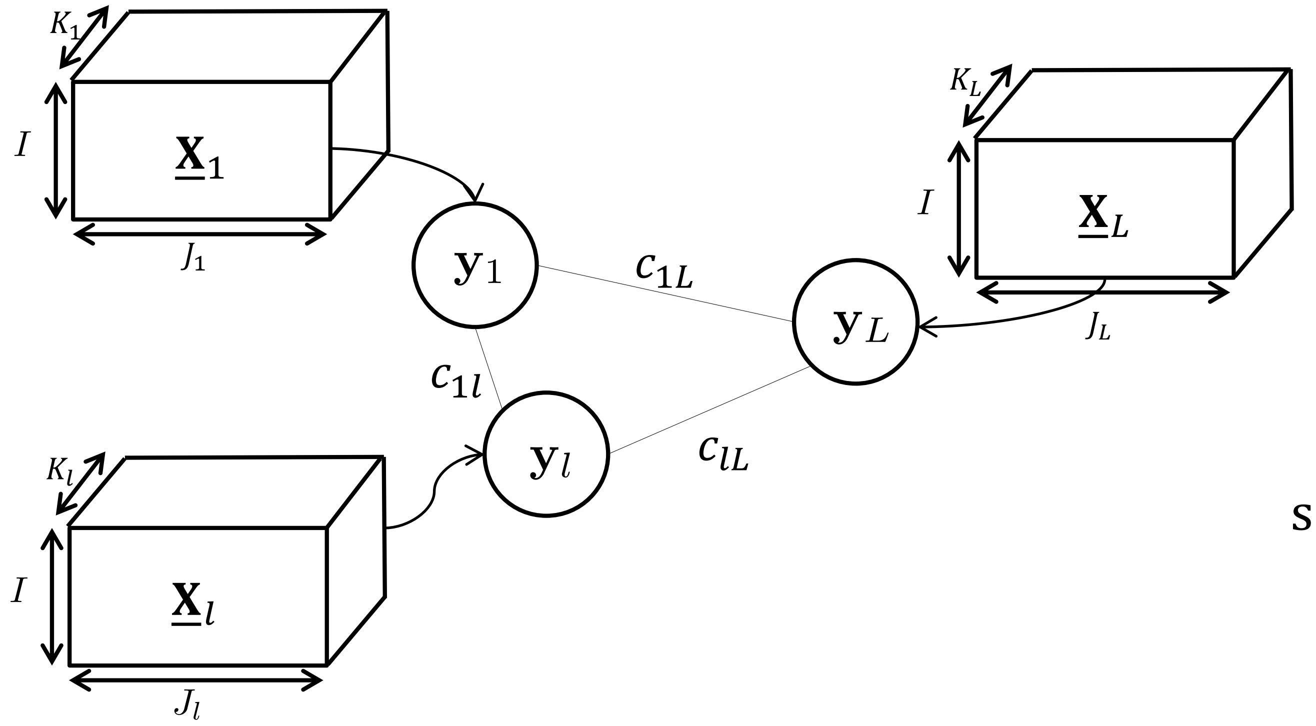
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad\qquad , l = 1, \ldots, L. \\ \end{cases}$$

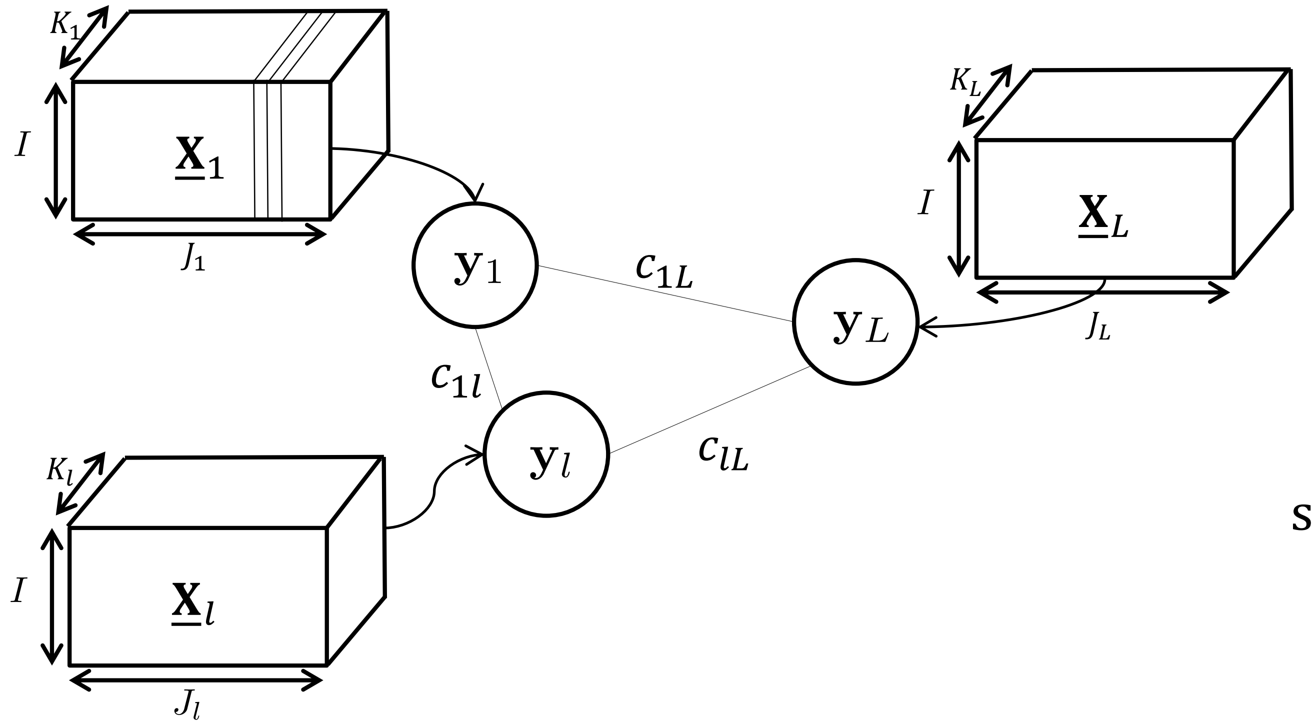
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad , l = 1, \ldots, L. \\ \end{cases}$$

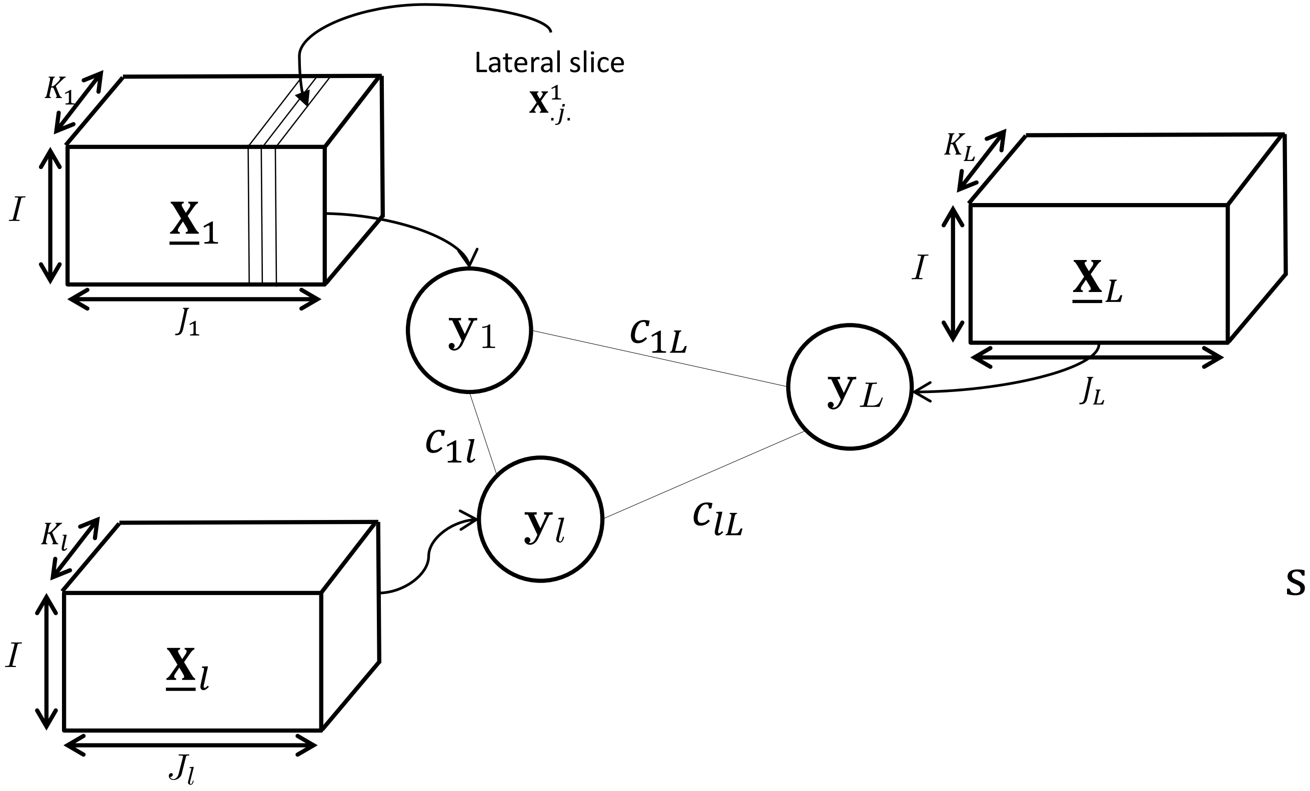
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \quad , l = 1, \ldots, L. \\ \end{cases}$$

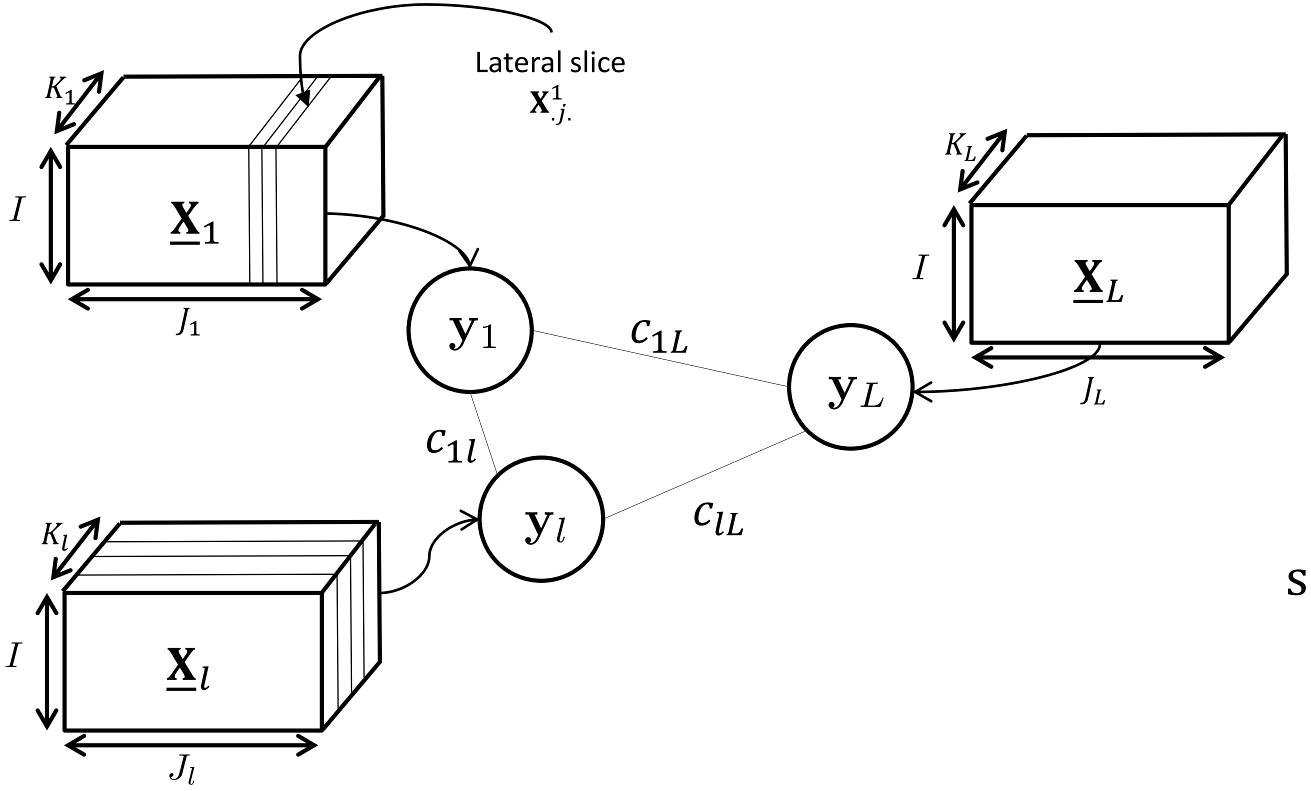
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad , l = 1, \ldots, L. \end{cases}$$

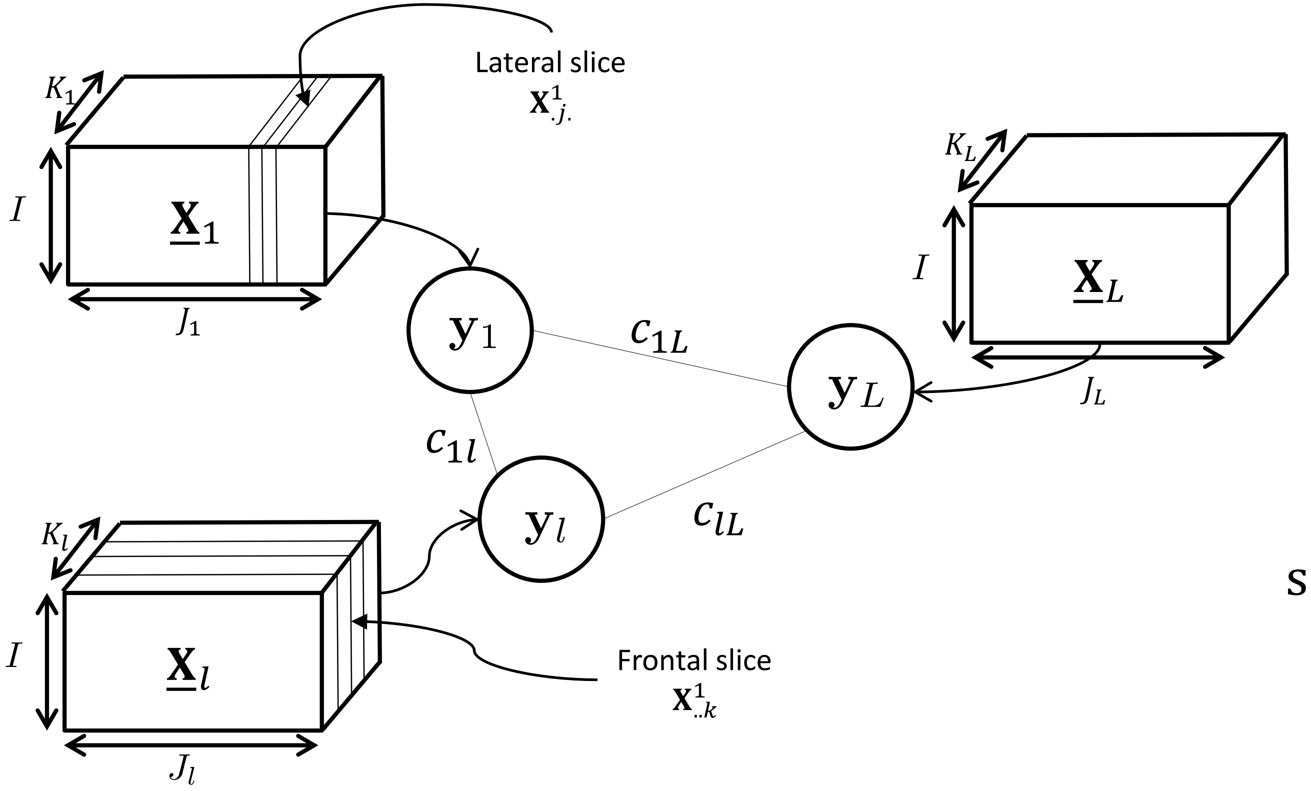
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad , l = 1, \ldots, L. \end{cases}$$

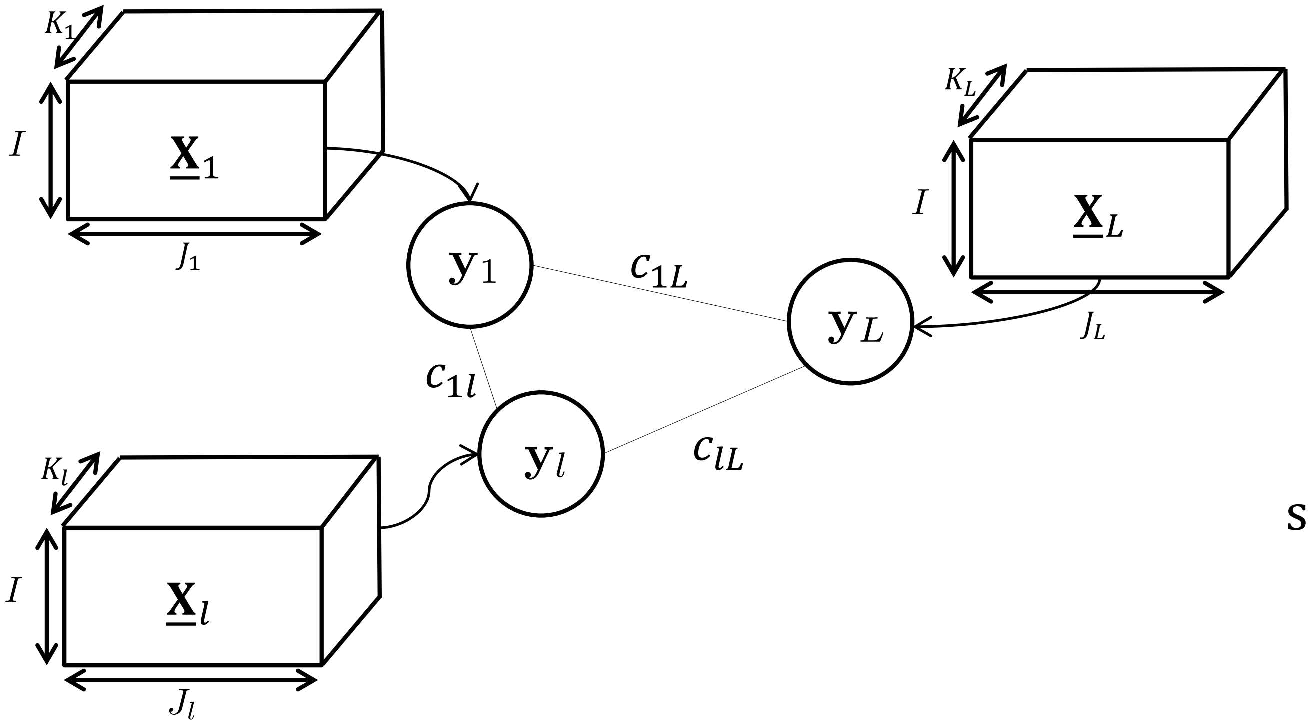
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\underline{\mathbf{X}}_k \mathbf{w}_k, \underline{\mathbf{X}}_l \mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad\qquad , l = 1, \ldots, L. \end{cases}$$

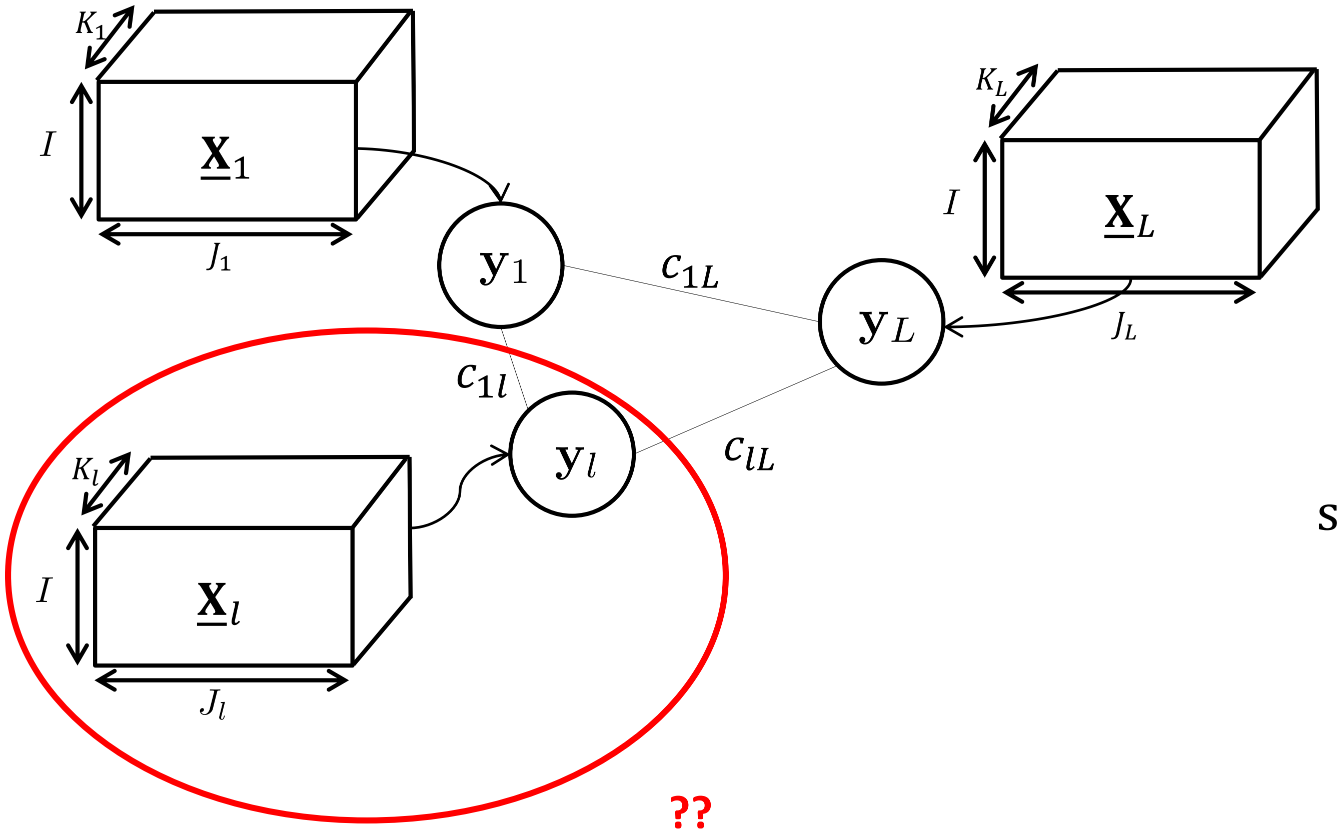
From RGCCA to Multiway GCCA



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad\qquad\qquad\qquad , l = 1, \ldots, L. \\ \end{cases}$$

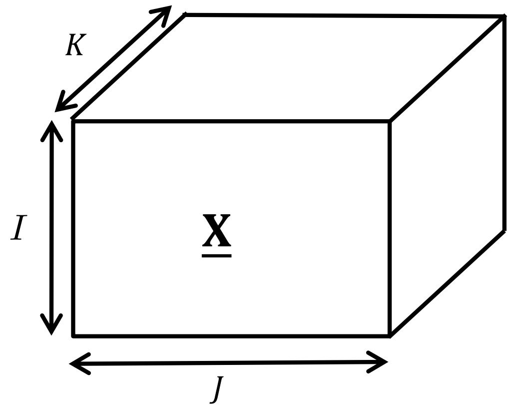
From RGCCA to Multiway GCCA



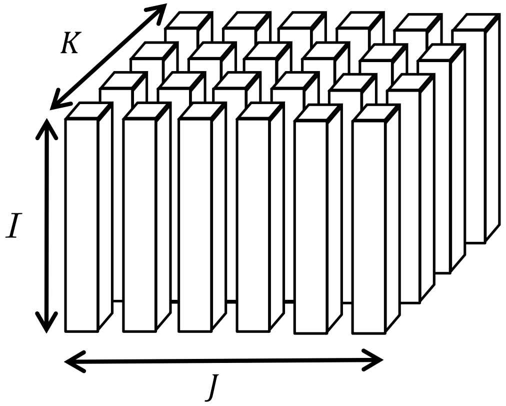
$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \qquad , l = 1, \ldots, L. \\ \end{cases}$$

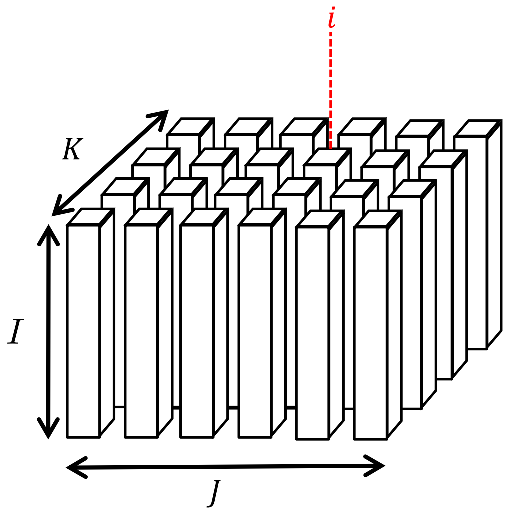
How Multiway is handled ?



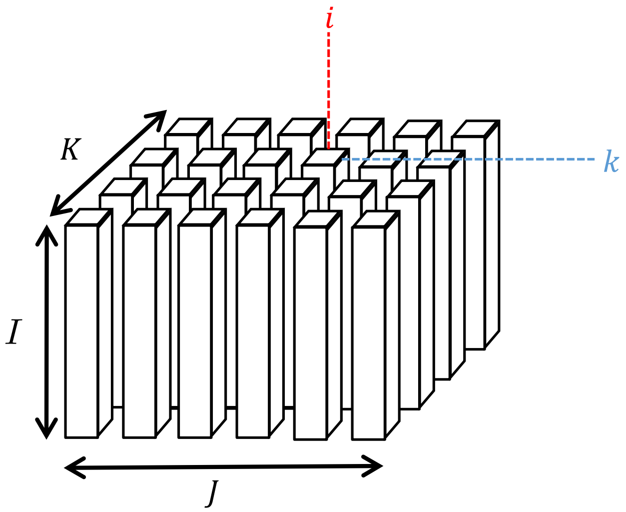
How Multiway is handled ?



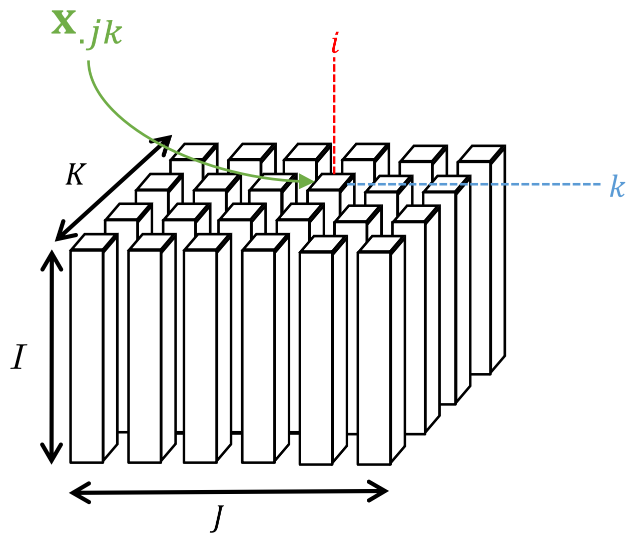
How Multiway is handled ?



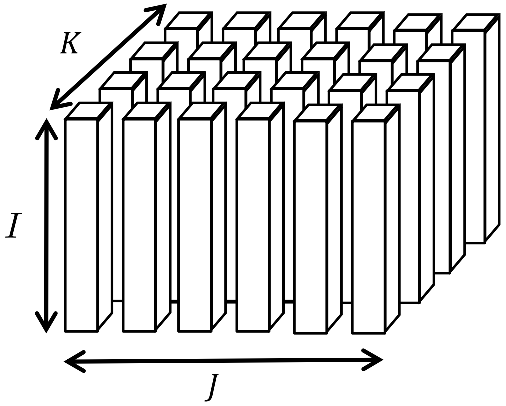
How Multiway is handled ?



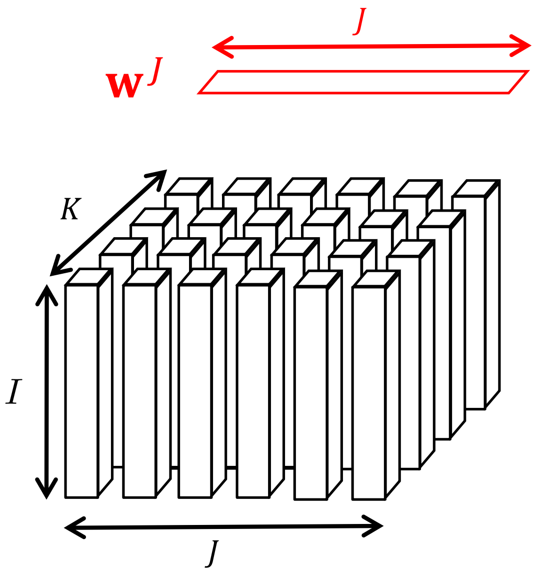
How Multiway is handled ?



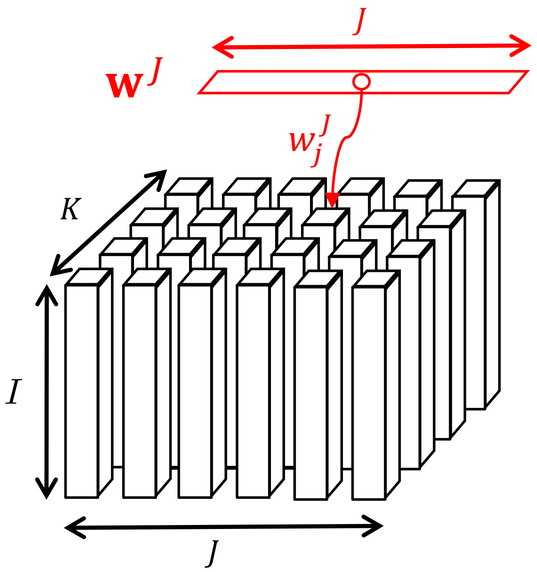
How Multiway is handled ?



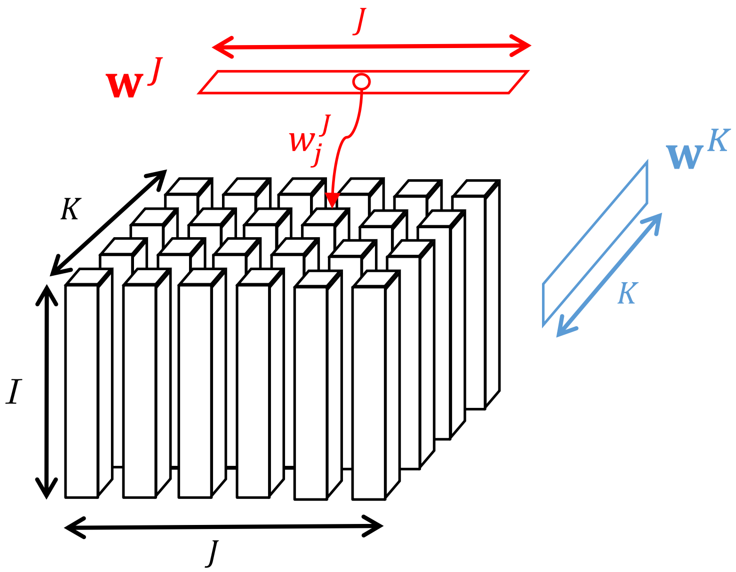
How Multiway is handled ?



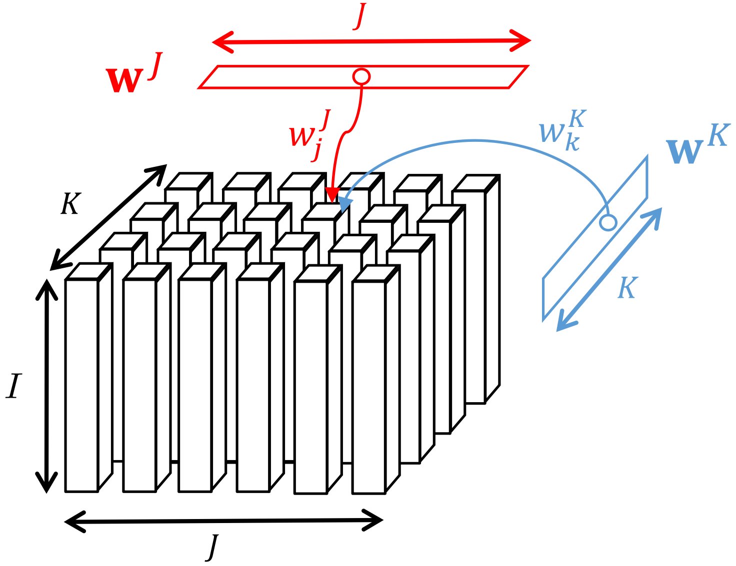
How Multiway is handled ?



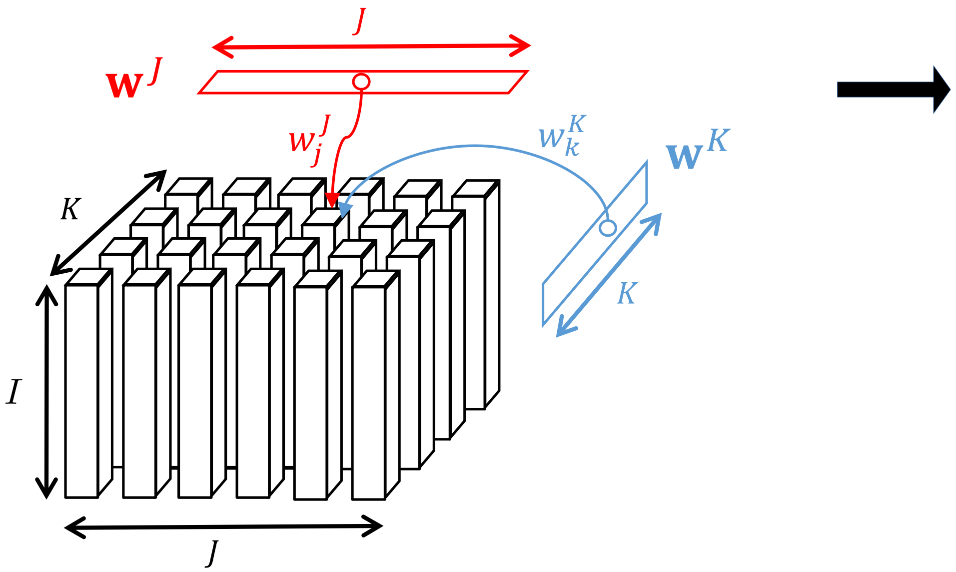
How Multiway is handled ?



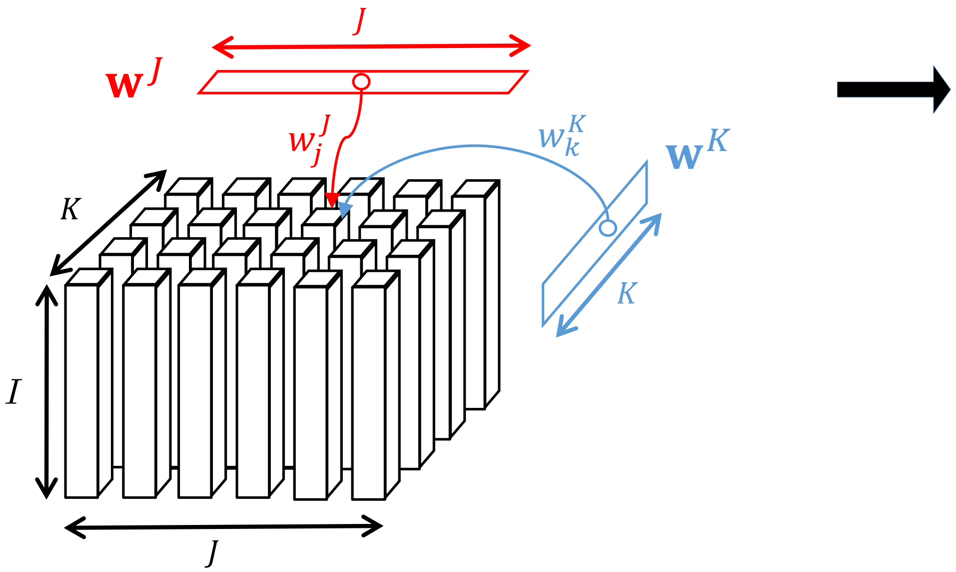
How Multiway is handled ?



How Multiway is handled ?

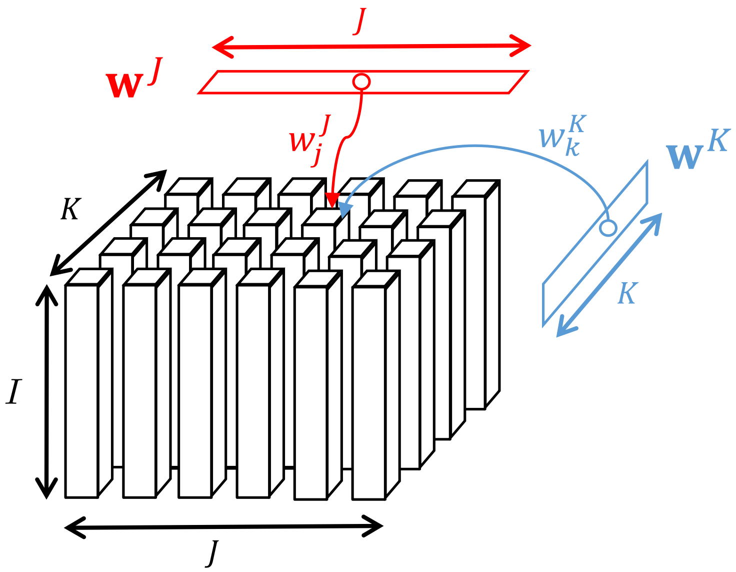


How Multiway is handled ?



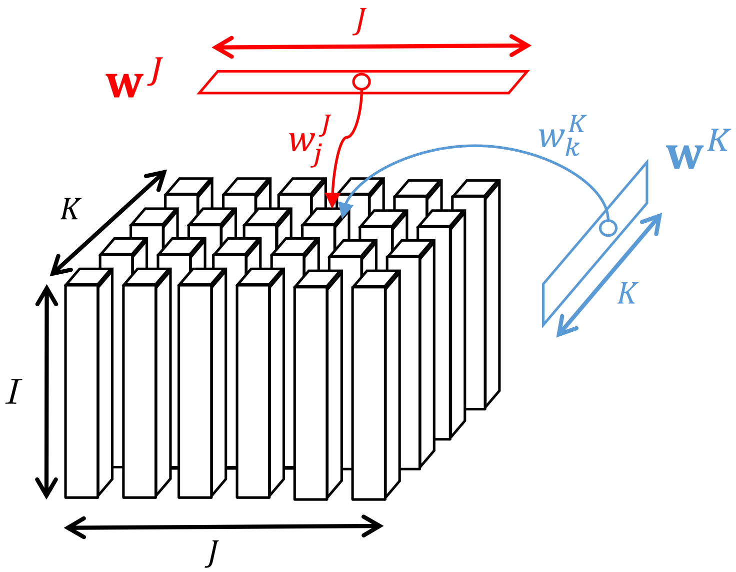
$$\mathbf{y} = \sum_{j=1}^{J} \sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

How Multiway is handled ?



$$\mathbf{y} = \sum_{j=1}^{J} \sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

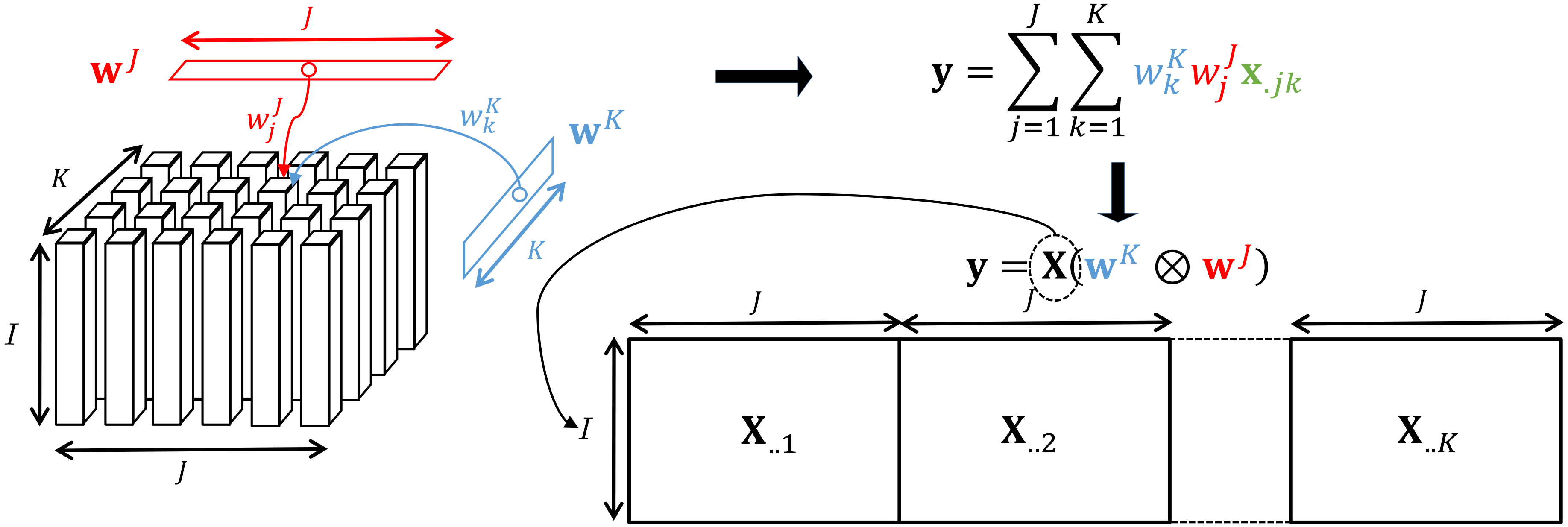
How Multiway is handled ?



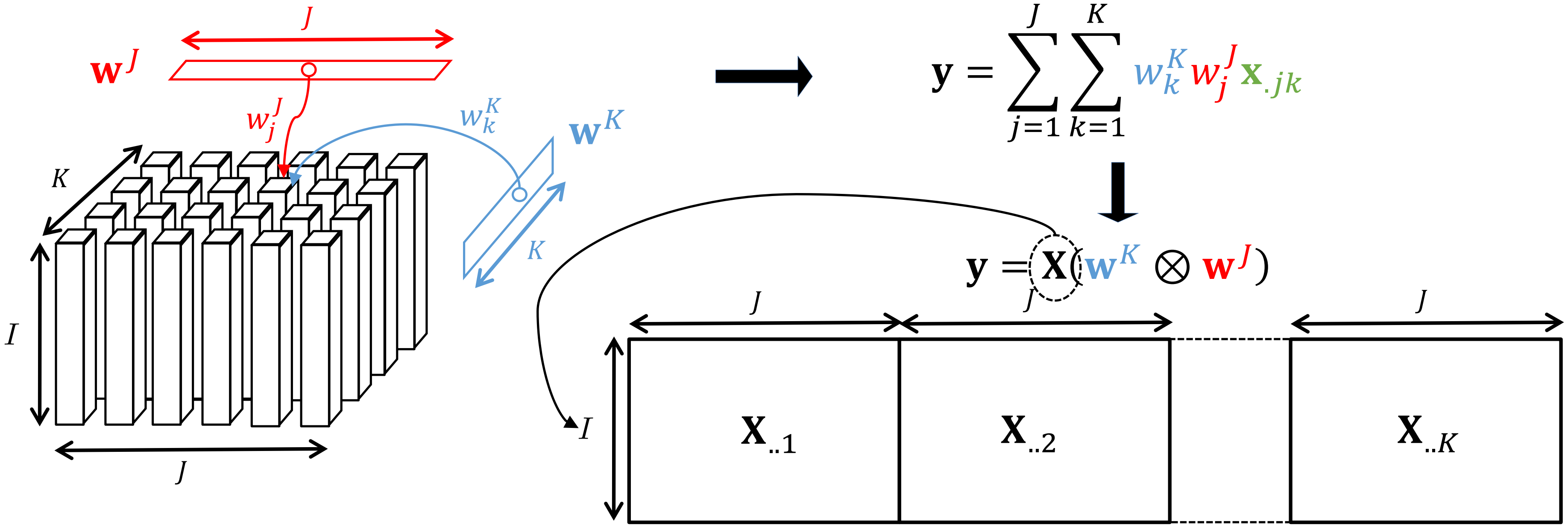
$$\mathbf{y} = \sum_{j=1}^{J} \sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

$$\mathbf{y} = \mathbf{X}(\mathbf{w}^K \otimes \mathbf{w}^J)$$

How Multiway is handled ?

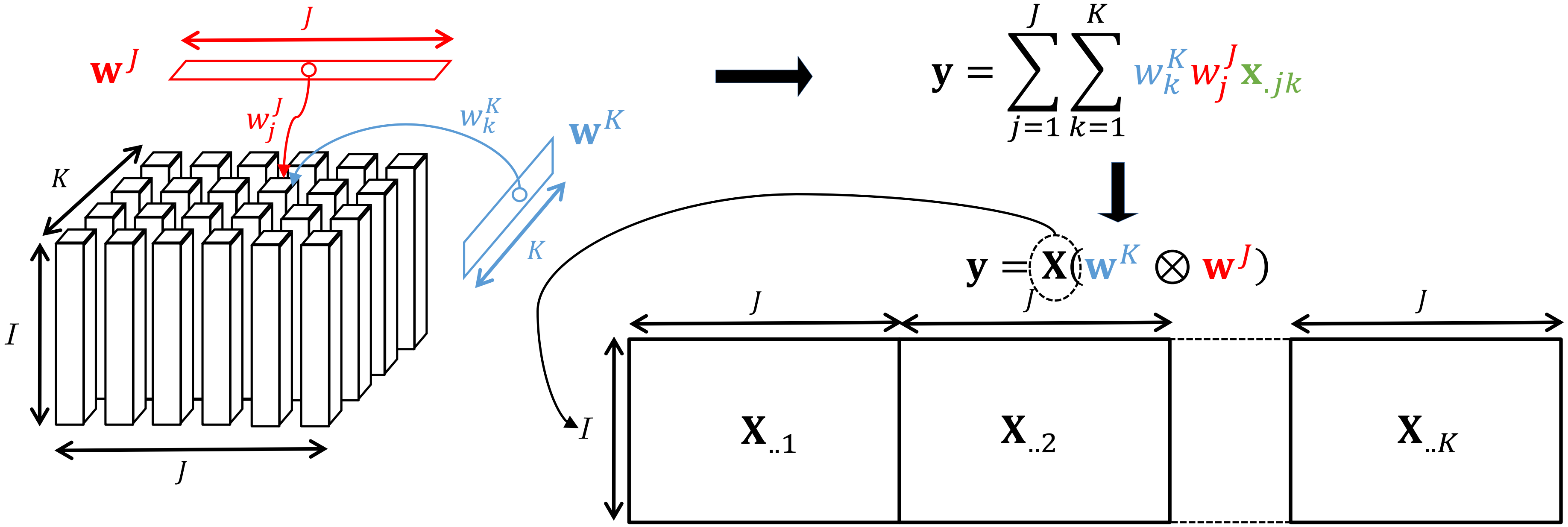


How Multiway is handled ?



Interest in taking into account 3-way structure with the Kronecker product:

How Multiway is handled ?



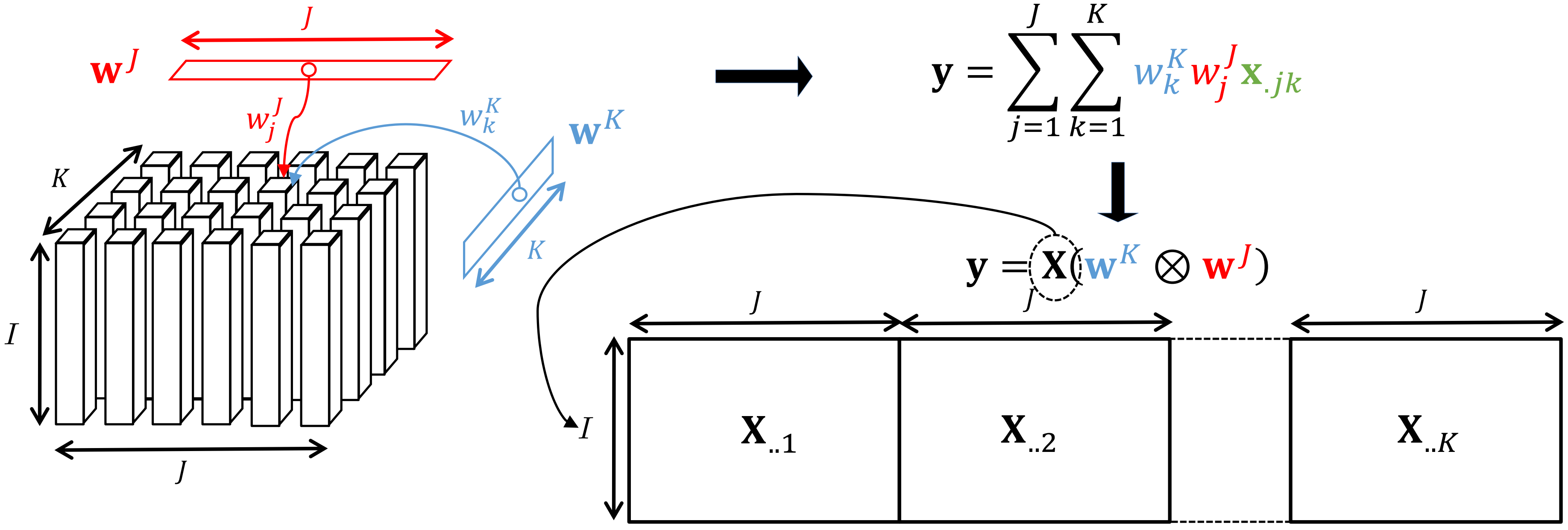
$$\mathbf{y} = \sum_{j=1}^{J}\sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

$$\mathbf{y} = \mathbf{X}(\mathbf{w}^K \otimes \mathbf{w}^J)$$

Interest in taking into account 3-way structure with the Kronecker product:

❖ Gain in interpretability thanks to vector weights specific to each dimension.

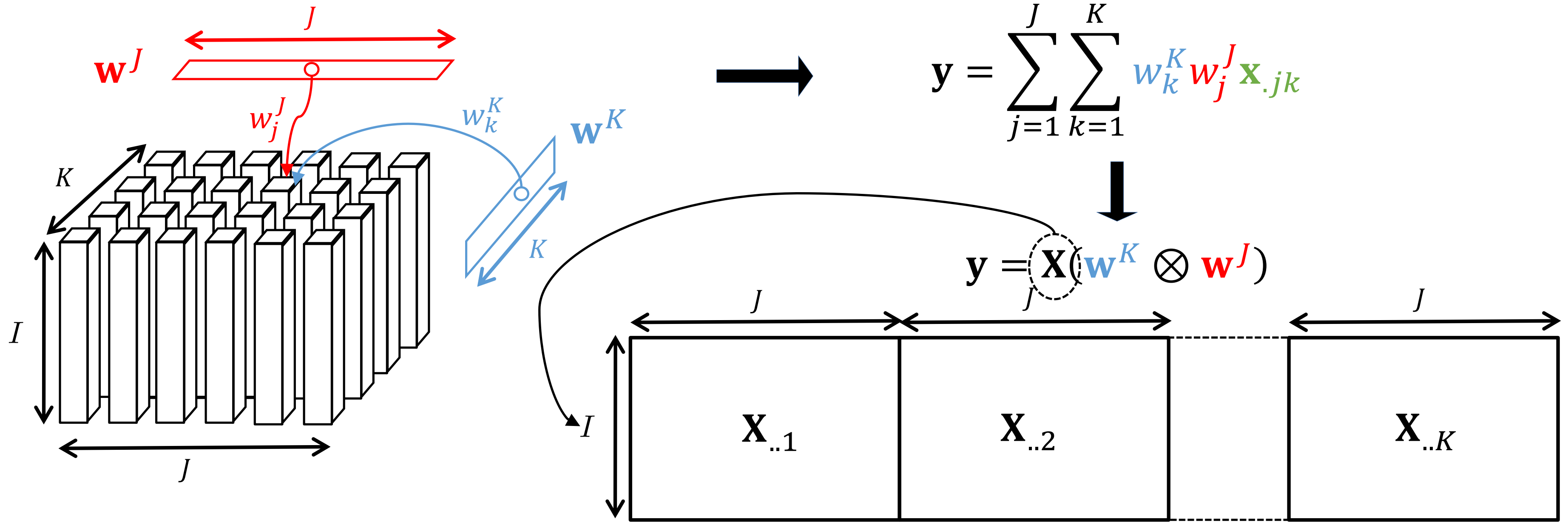
How Multiway is handled ?



Interest in taking into account 3-way structure with the Kronecker product:

- ❖ Gain in interpretability thanks to vector weights specific to each dimension.
- ❖ Less weights to estimate: from $J \times K$ to $J + K$.

How Multiway is handled ?

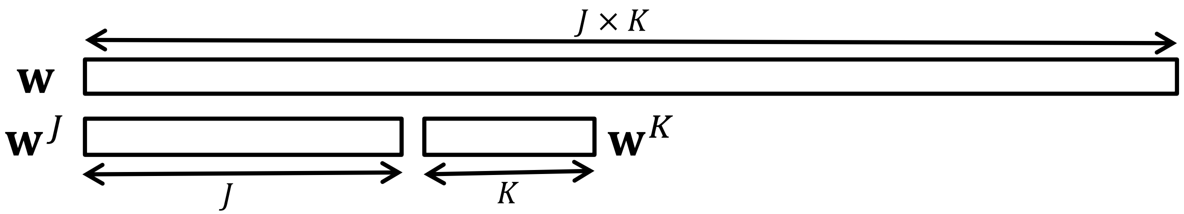


$$\mathbf{y} = \sum_{j=1}^{J} \sum_{k=1}^{K} w_k^K w_j^J \mathbf{x}_{.jk}$$

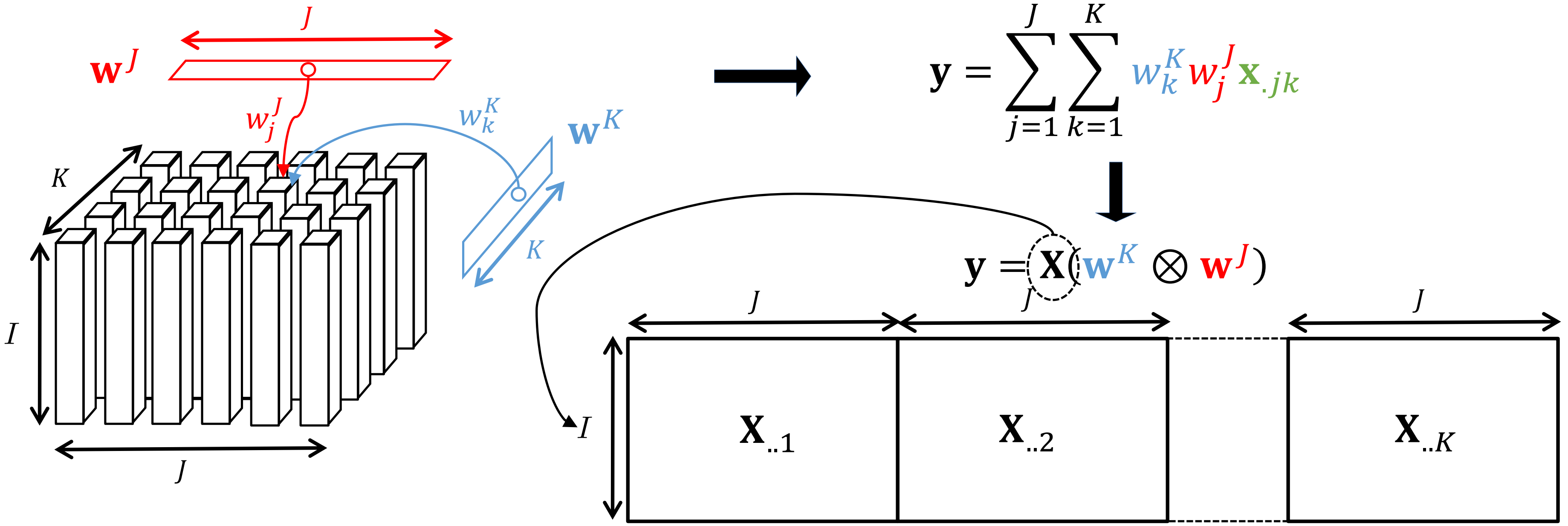
$$\mathbf{y} = \mathbf{X}(\mathbf{w}^K \otimes \mathbf{w}^J)$$

Interest in taking into account 3-way structure with the Kronecker product:

- ❖ Gain in interpretability thanks to vector weights specific to each dimension.
- ❖ Less weights to estimate: from $J \times K$ to $J + K$.

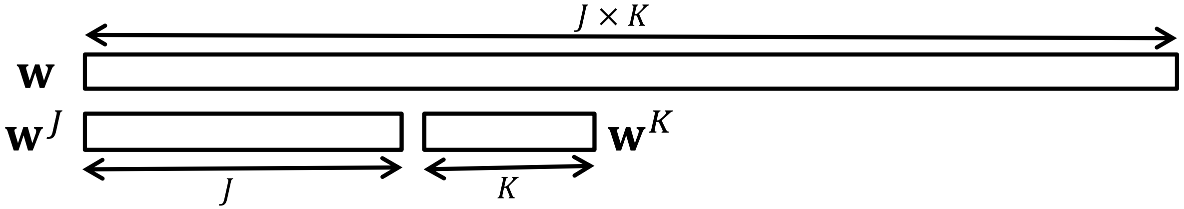


How Multiway is handled ?



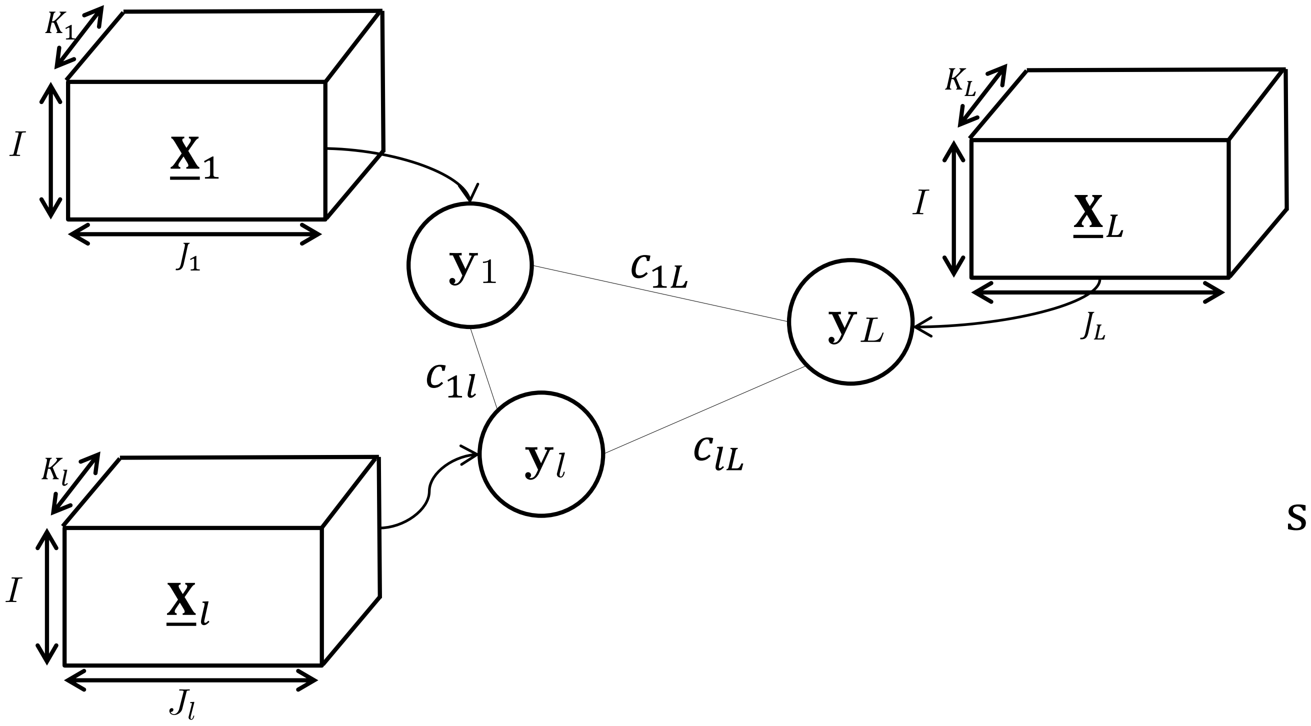
Interest in taking into account 3-way structure with the Kronecker product:

- Gain in interpretability thanks to vector weights specific to each dimension.
- Less weights to estimate: from $J \times K$ to $J + K$.



$J \times K$ Vs $J + K$

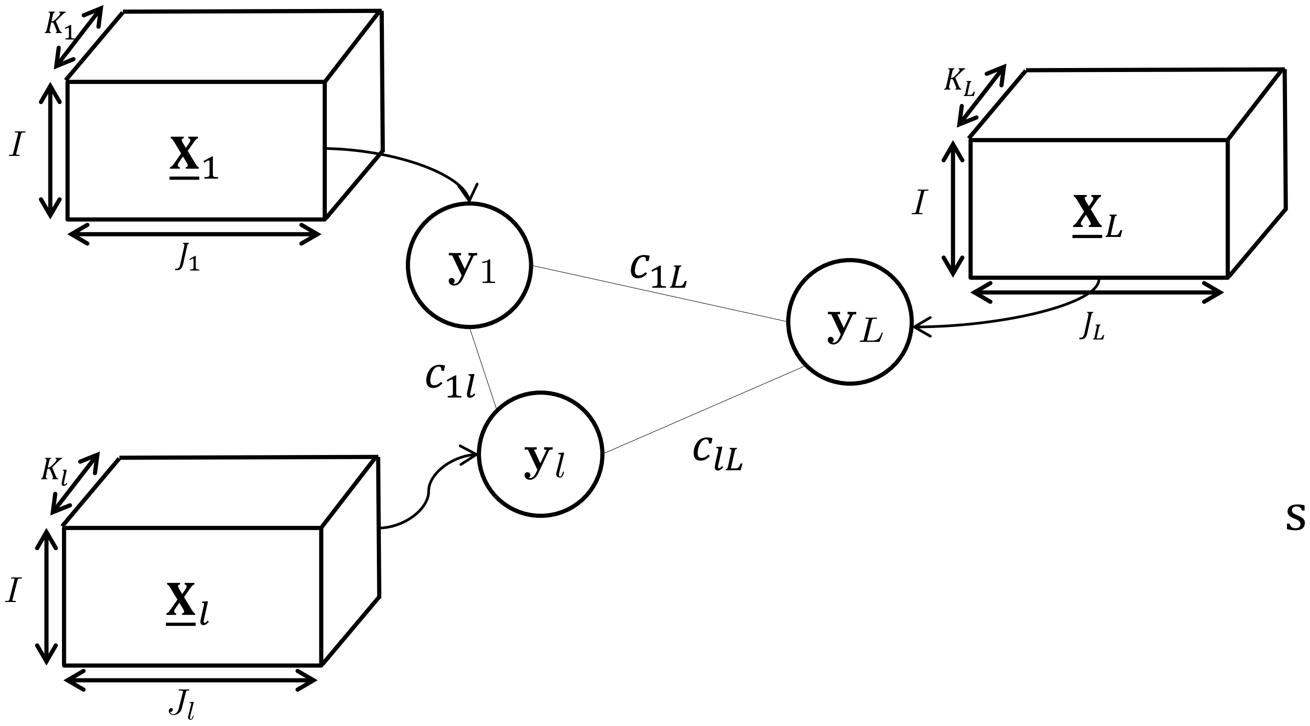
Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \quad , l = 1, \ldots, L. \\ \end{cases}$$

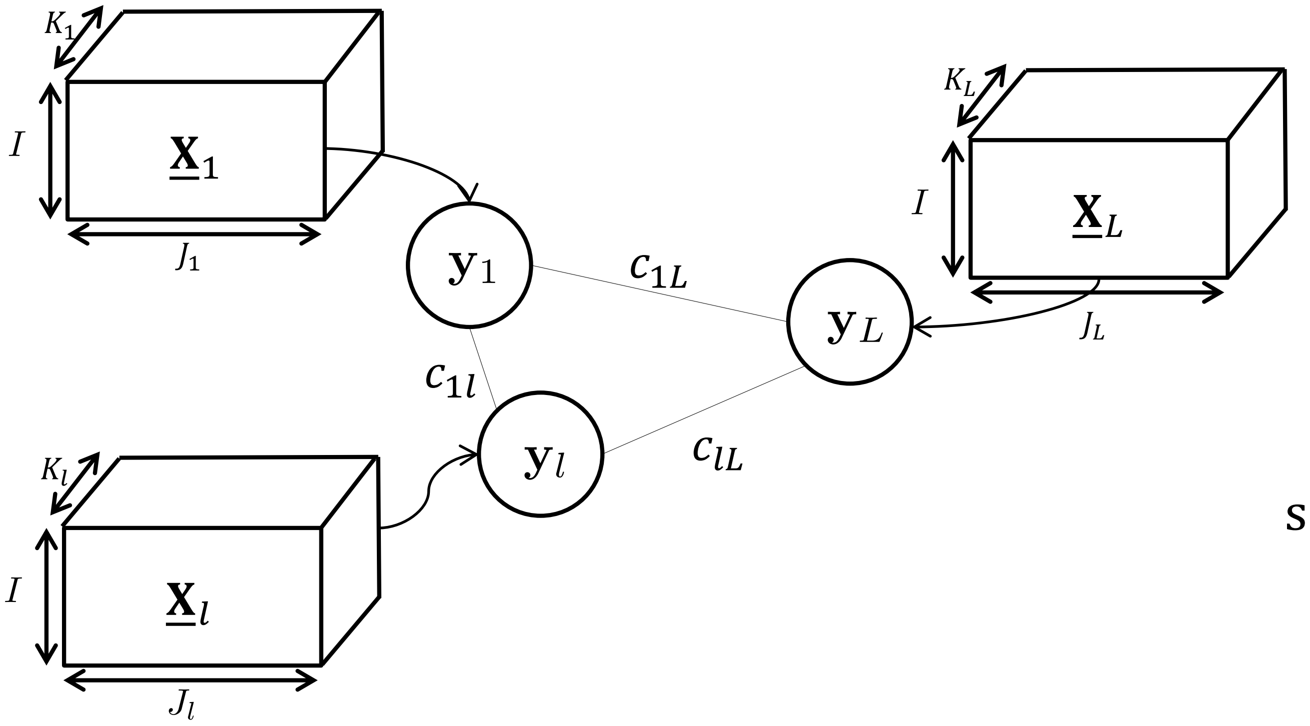
Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

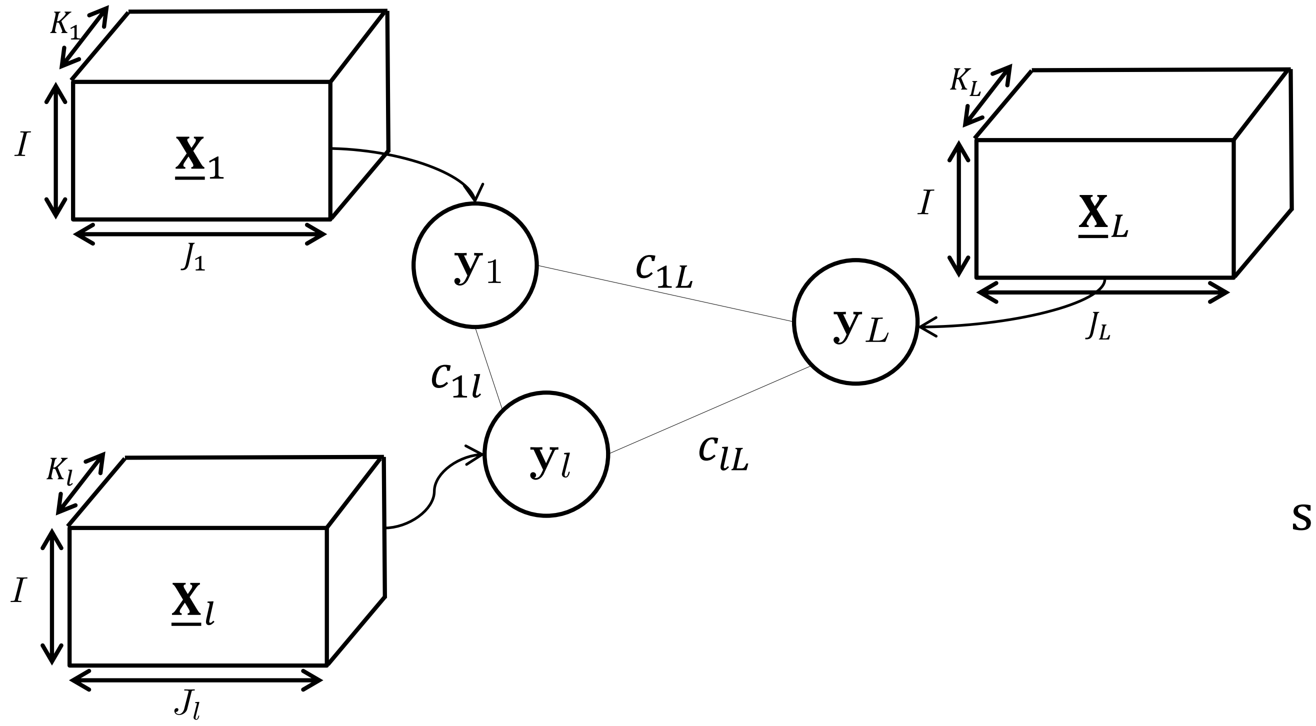
Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

Multiway Generalized Canonical Correlation Analysis (MGCCA)

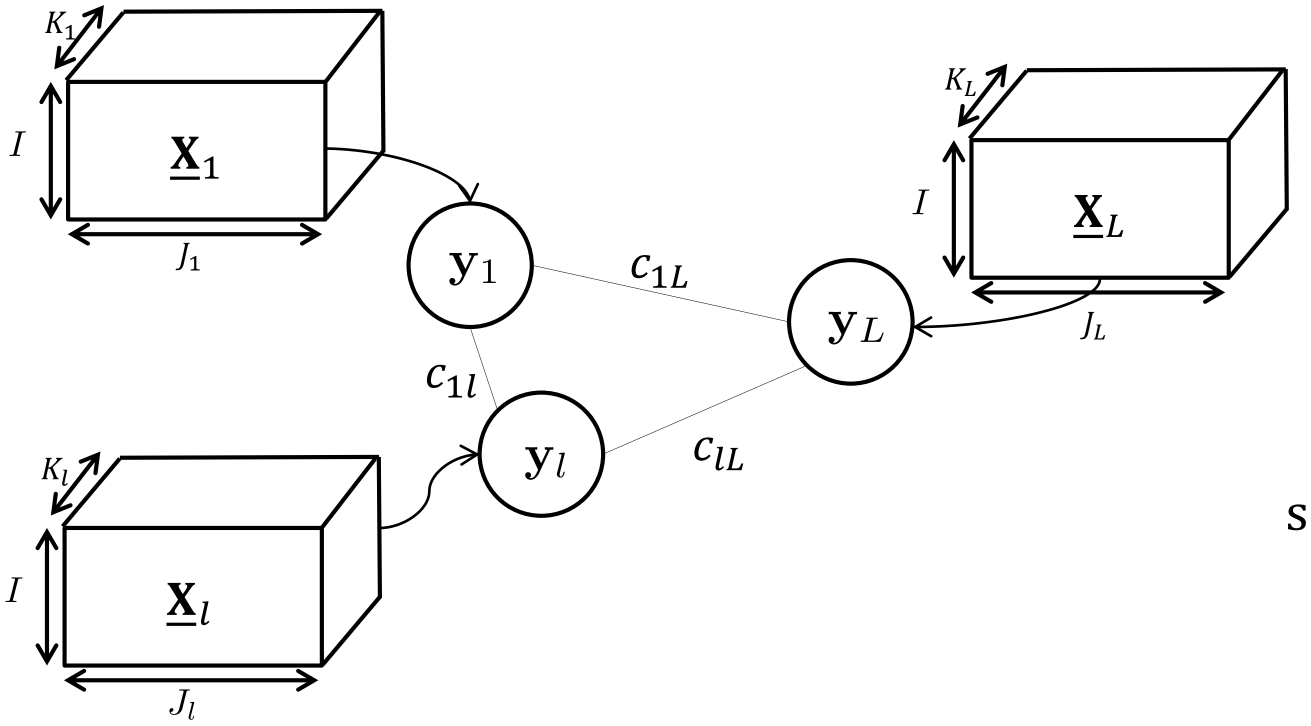


$$\max_{\mathbf{w}_1,\dots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \dots, L.$$

Example of such data: Electro-EncephaloGrams.

Multiway Generalized Canonical Correlation Analysis (MGCCA)

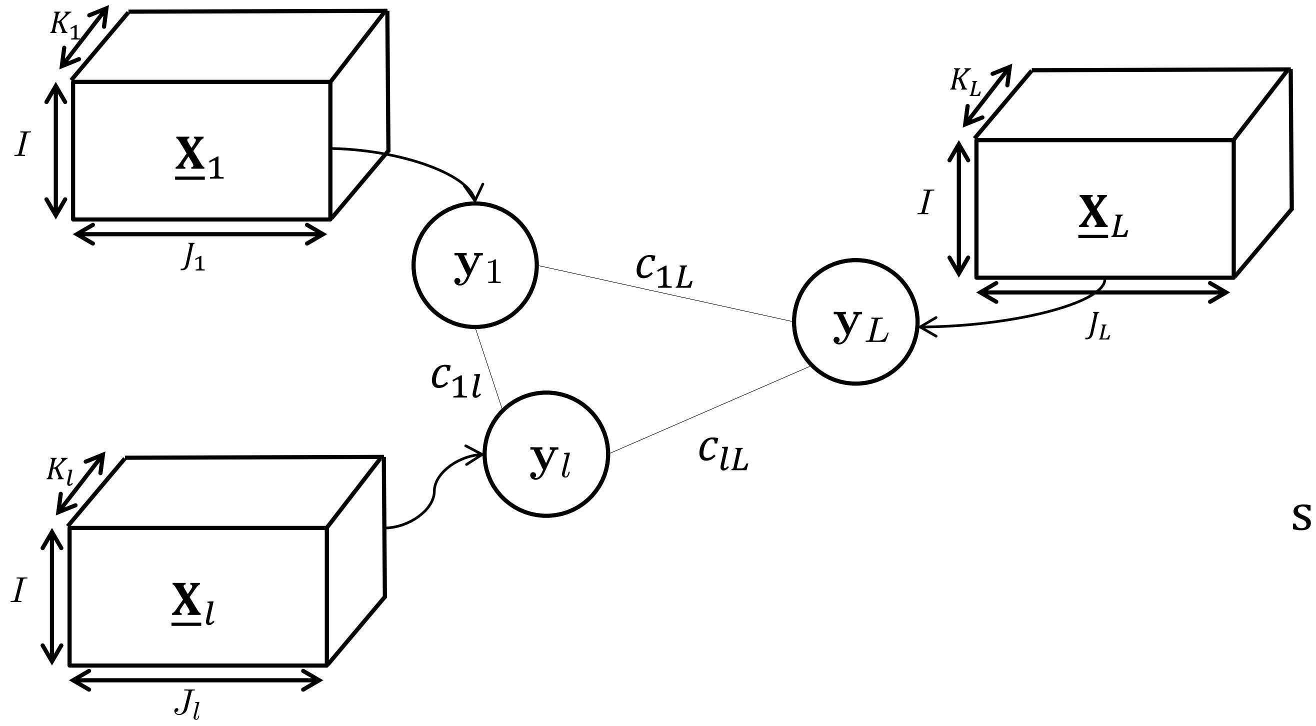


$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

Example of such data: Electro-EncephaloGrams.
Idea of the Algorithm:

Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

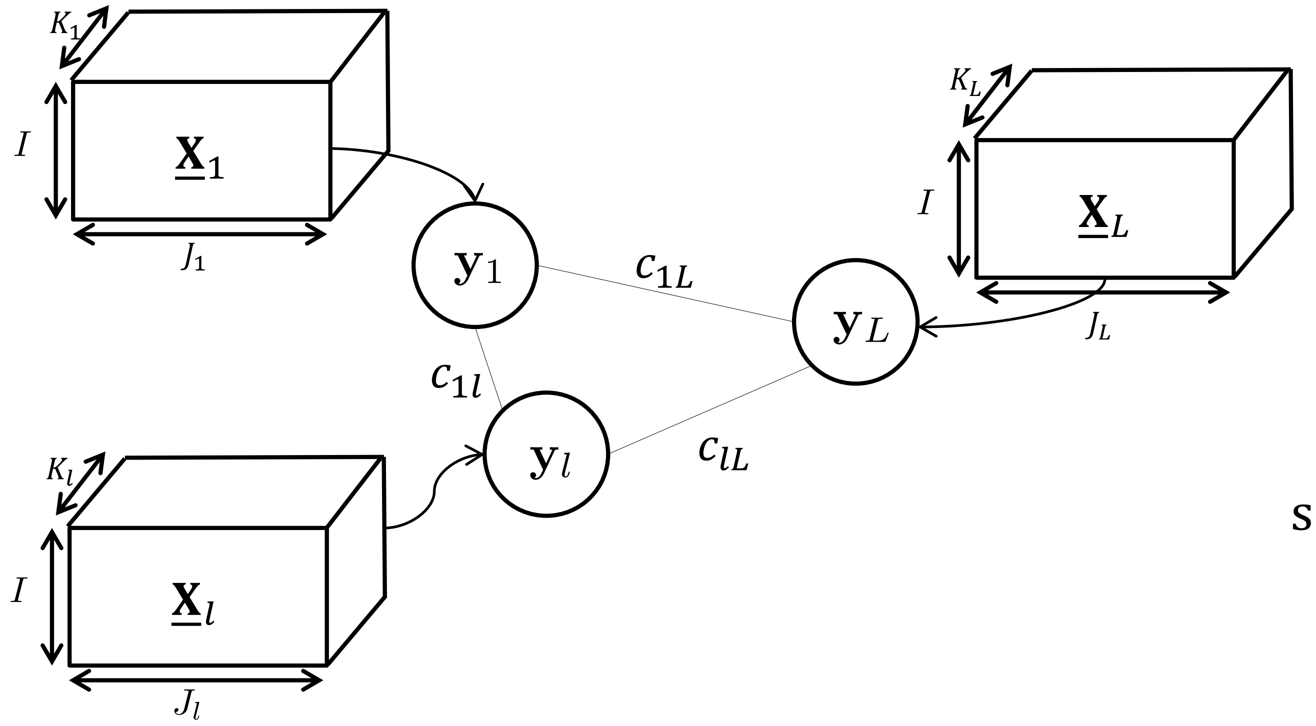
$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

Example of such data: Electro-EncephaloGrams.

Idea of the Algorithm:

1. Block Coordinate Ascent (BCA).

Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

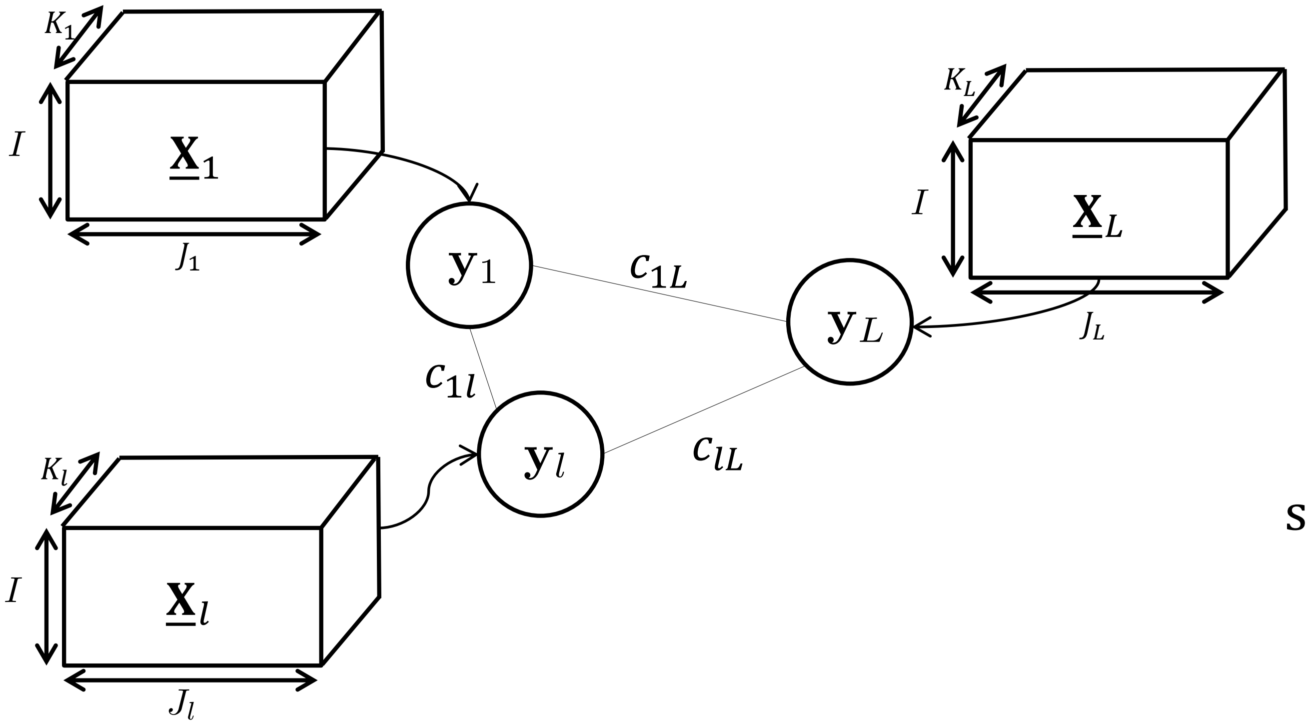
$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

Example of such data: Electro-EncephaloGrams.

Idea of the Algorithm:

1. Block Coordinate Ascent (BCA).
2. MM principle: each update is a SVD of a specic matrix of size $K_l \times J_l$.

Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

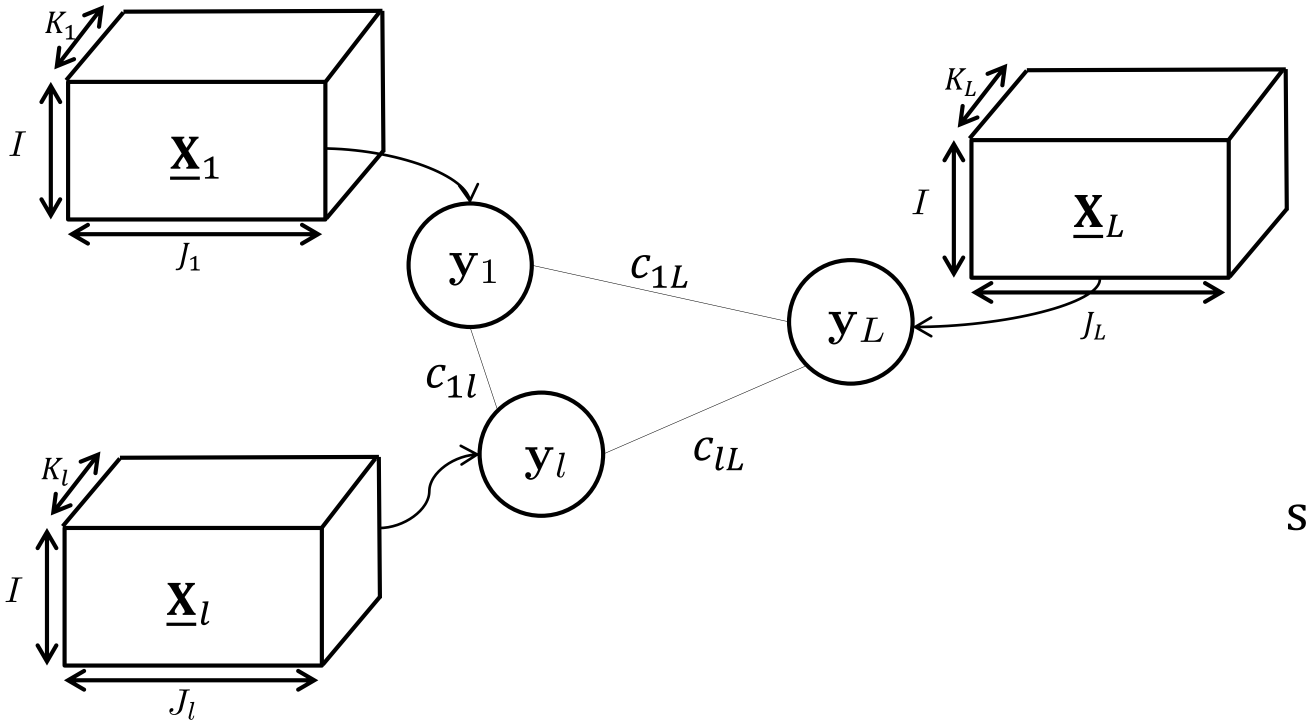
$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases} , l = 1, \ldots, L.$$

Example of such data: Electro-EncephaloGrams.
Idea of the Algorithm:

1. Block Coordinate Ascent (BCA).
2. MM principle: each update is a SVD of a specic matrix of size $K_l \times J_l$.

Global convergence of this algorithm was shown.

Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\max_{\mathbf{w}_1,\ldots,\mathbf{w}_L} \sum_{k,l=1}^{L} c_{kl}\, g\big(\mathrm{Cov}(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l)\big)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \ldots, L.$$

Example of such data: Electro-EncephaloGrams.

Idea of the Algorithm:

1. Block Coordinate Ascent (BCA).
2. MM principle: each update is a SVD of a specic matrix of size $K_l \times J_l$.

New extension with Tensor GCCA

Global convergence of this algorithm was shown.

		Core Optimization Problem			
		$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 \leq s_l \end{cases}$	SGCCA[3]			

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014)

RGCCA framework - State of the Art

		Core Optimization Problem			
		$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$				

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014)

		Core Optimization Problem			
		$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \lVert \mathbf{w}_l \rVert_2^2 = 1 \\ \lVert \mathbf{w}_l \rVert_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]			

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)

5. (Girka et al., 2024)

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

1. Introduction of the case study
2. **Unsupervised analysis with one-block:** Principal Component Analysis (PCA)
3. **Unsupervised analysis with two-blocks:**
   Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA)
4. **Unsupervised analysis with L-blocks:**
   Regularized Generalized Canonical Correlation Analysis (RGCCA)
5. **Supervised analysis with RGCCA**
6. **Variable selection in RGCCA:**
   Sparse Generalized Canonical Correlation Analysis (SGCCA)
7. **The flexible Optimization Framework of RGCCA**
   - The general principal
   - Extension to multi-way analysis
   - From Sequential to Global

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1, \ldots, \mathbf{w}_L}{\text{argmax}} \sum_{k,l=1}^{L} c_{kl} \, g\left(\text{Cov}\left(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l\right)\right)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1 \end{cases}, l = 1, \ldots, L.$$

Where:

- $\mathbf{w}_l \in \mathbb{R}^{J_l}$ is a block-weight vector.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1^{(1)},\ldots,\mathbf{w}_L^{(1)}}{\text{argmax}} \sum_{k,l=1}^{L} c_{kl} \quad g\left(\text{Cov}\left(\mathbf{X}_k\mathbf{w}_k^{(1)}, \mathbf{X}_l\mathbf{w}_l^{(1)}\right)\right)$$

$$\text{s.t.} \begin{cases} {\mathbf{w}_l^{(1)}}^\top \mathbf{M}_l \mathbf{w}_l^{(1)} = 1 \\ \end{cases} \quad , l = 1, \ldots, L.$$

Where:

- ❖ $\mathbf{w}_l^{(1)} \in \mathbb{R}^{J_l}$ is a the first block-weight vector.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1^{(2)},\ldots,\mathbf{w}_L^{(2)}}{\text{argmax}} \sum_{k,l=1}^{L} c_{kl} \quad g\left(\text{Cov}\left(\mathbf{X}_k\mathbf{w}_k^{(2)}, \mathbf{X}_l\mathbf{w}_l^{(2)}\right)\right)$$

$$\text{s.t.} \begin{cases} {\mathbf{w}_l^{(2)}}^\top \mathbf{M}_l \mathbf{w}_l^{(2)} = 1 \\ {\mathbf{y}_l^{(1)}}^\top \mathbf{X}_l \mathbf{w}_l^{(2)} = 0 \end{cases} \quad , l = 1, \ldots, L.$$

Where:

- $\mathbf{w}_l^{(1)} \in \mathbb{R}^{J_l}$ is a the first block-weight vector.
- $\mathbf{w}_l^{(2)} \in \mathbb{R}^{J_l}$is a the second block−weight vector.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1^{(2)},\ldots,\mathbf{w}_L^{(2)}}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \quad g\left(\operatorname{Cov}\left(\mathbf{X}_k\mathbf{w}_k^{(2)}, \mathbf{X}_l\mathbf{w}_l^{(2)}\right)\right)$$

$$\text{s.t.} \begin{cases} \mathbf{w}_l^{(2)\top}\mathbf{M}_l\mathbf{w}_l^{(2)} = 1 \\ \mathbf{y}_l^{(1)\top}\mathbf{X}_l\mathbf{w}_l^{(2)} = 0 \end{cases} \quad , l = 1, \ldots, L.$$

Where:

- ❖ $\mathbf{w}_l^{(1)} \in \mathbb{R}^{J_l}$ is a the first block-weight vector.
- ❖ $\mathbf{w}_l^{(2)} \in \mathbb{R}^{J_l}$is a the second block−weight vector.
- ❖ …

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1^{(1)},\ldots,\mathbf{w}_L^{(1)}}{\text{argmax}} \sum_{k,l=1}^{L} c_{kl} \quad g\left(\text{Cov}\left(\mathbf{X}_k\mathbf{w}_k^{(1)}, \mathbf{X}_l\mathbf{w}_l^{(1)}\right)\right)$$

Where:

❖ $\mathbf{w}_l^{(1)} \in \mathbb{R}^{J_l}$is a the first block−weight vector.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{w}_1^{(r)},\ldots,\mathbf{w}_L^{(r)}}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \sum_{r=1}^{R} g\left(\operatorname{Cov}\left(\mathbf{X}_k\mathbf{w}_k^{(r)}, \mathbf{X}_l\mathbf{w}_l^{(r)}\right)\right)$$

Where:

- $\mathbf{w}_l^{(\mathrm{r})} \in \mathbb{R}^{J_l}$ is a the r^{th} block–weight vector.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{W}_1,\ldots,\mathbf{W}_L}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \operatorname{Trace}\left(g\left(\operatorname{Cov}(\mathbf{X}_k \mathbf{W}_k, \mathbf{X}_l \mathbf{W}_l)\right)\right)$$

Where:

- $\mathbf{w}_l^{(\mathrm{r})} \in \mathbb{R}^{J_l}$ is a the r^{th} block−weight vector.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \ldots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$ is a block−weight matrix.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{W}_1,\ldots,\mathbf{W}_L}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \operatorname{Trace}\Big(g\big(\operatorname{Cov}(\mathbf{X}_k\mathbf{W}_k, \mathbf{X}_l\mathbf{W}_l)\big)\Big)$$

$$\text{s.t.} \qquad \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R \qquad , l = 1, \ldots, L.$$

Where:

- ❖ $\mathbf{w}_l^{(\mathrm{r})} \in \mathbb{R}^{J_l}$ is a the r^{th} block–weight vector.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \ldots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$ is a block–weight matrix.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{W}_1,\ldots,\mathbf{W}_L}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \operatorname{Trace}\left(g\left(\operatorname{Cov}(\mathbf{X}_k\mathbf{W}_k, \mathbf{X}_l\mathbf{W}_l)\right)\right) \qquad f(\mathbf{W}_1, \ldots, \mathbf{W}_L)$$

$$\text{s.t.} \qquad \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R \qquad , l = 1, \ldots, L.$$

Where:

- ❖ $\mathbf{w}_l^{(\mathrm{r})} \in \mathbb{R}^{J_l}$ is a the r^{th} block−weight vector.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \ldots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$ is a block−weight matrix.

Optimization criterion: From Sequential to Global

$$\underset{\mathbf{W}_1,\ldots,\mathbf{W}_L}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} \operatorname{Trace}\left(g\left(\operatorname{Cov}(\mathbf{X}_k\mathbf{W}_k, \mathbf{X}_l\mathbf{W}_l)\right)\right) \qquad f\underbrace{(\mathbf{W}_1, \ldots, \mathbf{W}_L)}_{\mathbf{W}}$$

$$\text{s.t.} \qquad \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R \qquad , l = 1, \ldots, L.$$

Where:

- ❖ $\mathbf{w}_l^{(\mathrm{r})} \in \mathbb{R}^{J_l}$ is a the r^{th} block–weight vector.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \ldots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$ is a block–weight matrix.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \operatorname{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \operatorname{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \operatorname{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\text{argmax}} \quad \text{Trace}(\nabla_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\nabla_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- R the number of components to extract.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\text{argmax}} \quad \text{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- R the number of components to extract.

➡ <u>Closed form solution:</u> the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \operatorname{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- ❖ $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- ❖ R the number of components to extract.

➡ Closed form solution: the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

➡ Pros:

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\text{argmax}} \quad \text{Trace}(\nabla_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- ❖ $\nabla_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- ❖ R the number of components to extract.

Closed form solution: the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

Pros:

- ❖ A single optimization problem allows to extract all components simultaneously.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\text{argmax}} \quad \text{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- R the number of components to extract.

➡ Closed form solution: the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

➡ Pros:

- A single optimization problem allows to extract all components simultaneously.
- The obtain algorithm is rather simple (simple update) and is globally convergent.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\text{argmax}} \quad \text{Trace}(\nabla_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- ❖ $\nabla_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- ❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- ❖ R the number of components to extract.

Closed form solution: the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

Pros:

- ❖ A single optimization problem allows to extract all components simultaneously.
- ❖ The obtain algorithm is rather simple (simple update) and is globally convergent.
- ❖ It is possible now to add constraints across components.

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \mathrm{Trace}(\boldsymbol{\nabla}_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

❖ $\boldsymbol{\nabla}_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.

❖ $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.

❖ R the number of components to extract.

<u>Closed form solution:</u> the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

<u>Pros</u>:

❖ A single optimization problem allows to extract all components simultaneously.

❖ The obtain algorithm is rather simple (simple update) and is globally convergent.

❖ It is possible now to add constraints across components.

<u>Cons:</u>

Principle of the Global RGCCA Algorithm

Following the optimization framework of RGCCA, the core optimization problem is:

$$\underset{\mathbf{W}_l, \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R}{\operatorname{argmax}} \quad \mathrm{Trace}(\nabla_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$$

Where:

- $\nabla_l f$ is the partial derivate of f with respect to $\mathbf{W}_l$.
- $\mathbf{W}_l = \left[\mathbf{w}_l^{(1)}, \dots, \mathbf{w}_l^{(\mathrm{R})}\right] \in \mathbb{R}^{J_l \times R}$.
- R the number of components to extract.

Closed form solution: the rank-R Singular Value Decomposition (SVD) of a specic matrix of dimension $J_l \times R$.

Pros:

- A single optimization problem allows to extract all components simultaneously.
- The obtain algorithm is rather simple (simple update) and is globally convergent.
- It is possible now to add constraints across components.

Cons:

- In this form, we have to extract the same number of component per block.

RGCCA framework - State of the Art

		Core Optimization Problem			
		$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \lVert \mathbf{w}_l \rVert_2^2 = 1 \\ \lVert \mathbf{w}_l \rVert_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]			

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)

RGCCA framework - State of the Art

		Core Optimization Problem			
		Sequential	Global		
		$\max \nabla_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$			
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]			Constraints
	$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]			

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

RGCCA framework - State of the Art

Constraints	Core Optimization Problem: Sequential	Core Optimization Problem: Global	Constraints
	$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W}^{\mathbf{s}})^\top \mathbf{W}_l)$	
$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]		
$\begin{cases} \Vert\mathbf{w}_l\Vert_2^2 = 1 \\ \Vert\mathbf{w}_l\Vert_1 \leq s_l \end{cases}$	SGCCA[3]		
$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]		

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)

INSTITUT FRANÇAIS DE BIOINFORMATIQUE
elixir FRANCE

RGCCA framework - State of the Art

		Core Optimization Problem			
		Sequential	Global		
		$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W^s})^\top \mathbf{W}_l)$		
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]		$\mathbf{W}_l \in \Omega_l$	Constraints
	$\begin{cases} \Vert\mathbf{w}_l\Vert_2^2 = 1 \\ \Vert\mathbf{w}_l\Vert_1 \le s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]			

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)

RGCCA framework - State of the Art

		Core Optimization Problem			
		Sequential	Global		
		$\max \nabla_l f(\mathbf{w^s})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\nabla_l f(\mathbf{W^s})^\top \mathbf{W}_l)$		
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$	Constraints
	$\begin{cases} \lVert\mathbf{w}_l\rVert_2^2 = 1 \\ \lVert\mathbf{w}_l\rVert_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]			

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024) 6. (Gloaguen, 2020) 7. (Girka, 2023)

RGCCA framework - State of the Art

	Core Optimization Problem		
	Sequential	Global	
	$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W}^{\mathbf{s}})^\top \mathbf{W}_l)$	
Constraints			**Constraints**
$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$
$\begin{cases} \Vert\mathbf{w}_l\Vert_2^2 = 1 \\ \Vert\mathbf{w}_l\Vert_1 \leq s_l \end{cases}$	SGCCA[3]		
$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]		$\begin{cases} \mathbf{W}_l = \mathbf{W}_l^K \odot \mathbf{W}_l^J \\ \mathbf{W}_l \in \Omega_l \end{cases}$

$$\omega_l = \left\{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\right\}$$

$$\Omega_l = \left\{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\right\}$$

1. (Tenenhaus and Tenenhaus, 2011)
2. (Tenenhaus, Tenenhaus and Groenen, 2017)
3. (Tenenhaus et al., 2014)
4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)
6. (Gloaguen, 2020)
7. (Girka, 2023)

RGCCA framework - State of the Art

		Core Optimization Problem			
		Sequential	Global		
		$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W}^{\mathbf{s}})^\top \mathbf{W}_l)$		
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$	Constraints
	$\begin{cases} \Vert\mathbf{w}_l\Vert_2^2 = 1 \\ \Vert\mathbf{w}_l\Vert_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]	Global MGCCA[6,7]	$\begin{cases} \mathbf{W}_l = \mathbf{W}_l^K \odot \mathbf{W}_l^J \\ \mathbf{W}_l \in \Omega_l \end{cases}$	

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011) 2. (Tenenhaus, Tenenhaus and Groenen, 2017) 3. (Tenenhaus et al., 2014) 4. (Gloaguen et al., 2022)
5. (Girka et al., 2024) 6. (Gloaguen, 2020) 7. (Girka, 2023)

RGCCA framework - State of the Art

		Core Optimization Problem			
		Sequential	Global		
		$\max \boldsymbol{\nabla}_l f(\mathbf{w^s})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W^s})^\top \mathbf{W}_l)$		
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$	Constraints
	$\begin{cases} \lVert\mathbf{w}_l\rVert_2^2 = 1 \\ \lVert\mathbf{w}_l\rVert_1 \leq s_l \end{cases}$	SGCCA[3]			
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]	Global MGCCA[6,7]	$\begin{cases} \mathbf{W}_l = \mathbf{W}_l^K \odot \mathbf{W}_l^J \\ \mathbf{W}_l \in \Omega_l \end{cases}$	
	Structured Sparsity				

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011)
2. (Tenenhaus, Tenenhaus and Groenen, 2017)
3. (Tenenhaus et al., 2014)
4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)
6. (Gloaguen, 2020)
7. (Girka, 2023)

RGCCA framework - State of the Art

		Core Optimization Problem		
		Sequential	Global	
		$\max \nabla_l f(\mathbf{w}^s)^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\nabla_l f(\mathbf{W}^s)^\top \mathbf{W}_l)$	
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$
	$\begin{cases} \lVert\mathbf{w}_l\rVert_2^2 = 1 \\ \lVert\mathbf{w}_l\rVert_1 \le s_l \end{cases}$	SGCCA[3]		
	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]	Global MGCCA[6,7]	$\begin{cases} \mathbf{W}_l = \mathbf{W}_l^K \odot \mathbf{W}_l^J \\ \mathbf{W}_l \in \Omega_l \end{cases}$
	Structured Sparsity	(i). Group-Lasso in the same framework[8] (ii). Other structured sparse penalties in other frameworks[6,9,10,11]		Constraints

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R}; \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011)
2. (Tenenhaus, Tenenhaus and Groenen, 2017)
3. (Tenenhaus et al., 2014)
4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)
6. (Gloaguen, 2020)
7. (Girka, 2023)
8. (Guillemot et al., 2021)
9. (Guigui et al., 2019)
10. (Chegraoui et al., 2023)
11. (Löfstedt et al., 2016)

RGCCA framework - State of the Art

Constraints		Core Optimization Problem			Constraints
		Sequential	**Global**		
		$\max \boldsymbol{\nabla}_l f(\mathbf{w}^{\mathbf{s}})^\top \mathbf{w}_l$	$\max \mathbf{Tr}(\boldsymbol{\nabla}_l f(\mathbf{W}^{\mathbf{s}})^\top \mathbf{W}_l)$		
Constraints	$\mathbf{w}_l \in \omega_l$	RGCCA[1,2]	Global RGCCA[6,7]	$\mathbf{W}_l \in \Omega_l$	Constraints
Constraints	$\begin{cases} \Vert\mathbf{w}_l\Vert_2^2 = 1 \\ \Vert\mathbf{w}_l\Vert_1 \leq s_l \end{cases}$	SGCCA[3]	In progress[7]	In progress[7]	Constraints
Constraints	$\begin{cases} \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \\ \mathbf{w}_l \in \omega_l \end{cases}$	MGCCA[4]/TGCCA[5]	Global MGCCA[6,7]	$\begin{cases} \mathbf{W}_l = \mathbf{W}_l^K \odot \mathbf{W}_l^J \\ \mathbf{W}_l \in \Omega_l \end{cases}$	Constraints
Constraints	Structured Sparsity	(i). Group-Lasso in the same framework[8] (ii). Other structured sparse penalties in other frameworks[6,9,10,11]	In progress[7]	In progress[7]	Constraints

$$\omega_l = \{\mathbf{w}_l \in \mathbb{R}^{J_l};\ \mathbf{w}_l^\top \mathbf{M}_l \mathbf{w}_l = 1\}$$

$$\Omega_l = \{\mathbf{W}_l \in \mathbb{R}^{J_l \times R};\ \mathbf{W}_l^\top \mathbf{M}_l \mathbf{W}_l = \mathbf{I}_R\}$$

1. (Tenenhaus and Tenenhaus, 2011)
2. (Tenenhaus, Tenenhaus and Groenen, 2017)
3. (Tenenhaus et al., 2014)
4. (Gloaguen et al., 2022)
5. (Girka et al., 2024)
6. (Gloaguen, 2020)
7. (Girka, 2023)
8. (Guillemot et al., 2021)
9. (Guigui et al., 2019)
10. (Chegraoui et al., 2023)
11. (Löfstedt et al., 2016)

RGCCA framework - State of the Art

RGCCA framework - State of the Art

Kernel GCCA (Tenenhaus, Philippe and Frouin, 2015):
In order to take estimate non-linear links between blocks.

RGCCA framework - State of the Art

Kernel GCCA (Tenenhaus, Philippe and Frouin, 2015):
In order to take estimate non-linear links between blocks.

Functional GCCA (Sort, Brusquet and Tenenhaus, 2023):
In order to handle longitudinal blocks.

RGCCA framework - State of the Art

Kernel GCCA (Tenenhaus, Philippe and Frouin, 2015):
In order to take estimate non-linear links between blocks.

Functional GCCA (Sort, Brusquet and Tenenhaus, 2023):
In order to handle longitudinal blocks.

Multi-group (Tenenhaus and Tenenhaus, 2014; Goujon, In progress):
Find relationships between variables within each group that are common to all groups.

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

RGCCA framework - State of the Art

Kernel GCCA (Tenenhaus, Philippe and Frouin, 2015):
In order to take estimate non-linear links between blocks.

Functional GCCA (Sort, Brusquet and Tenenhaus, 2023):
In order to handle longitudinal blocks.

Multi-group (Tenenhaus and Tenenhaus, 2014; Goujon, In progress):
Find relationships between variables within each group that are common to all groups.

Missing Values in RGCCA (Peltier et al., 2023).

RGCCA

https://github.com/rgcca-factory/RGCCA

RGCCA

The RGCCA framework is:

https://github.com/rgcca-factory/RGCCA

RGCCA

The RGCCA framework is:

❖ General as in encompasses a large number of methods in the multi-block literature.

https://github.com/rgcca-factory/RGCCA

RGCCA

<u>The RGCCA framework is:</u>

❖ General as in encompasses a large number of methods in the multi-block literature.
❖ Very flexible thanks to its optimization framework that easily allows to develop many extensions…

https://github.com/rgcca-factory/RGCCA

RGCCA

<u>The RGCCA framework is:</u>

- General as in encompasses a large number of methods in the multi-block literature.
- Very flexible thanks to its optimization framework that easily allows to develop many extensions…
- … and that also allows to combine these extensions.

https://github.com/rgcca-factory/RGCCA

RGCCA

The RGCCA framework is:

- General as in encompasses a large number of methods in the multi-block literature.
- Very flexible thanks to its optimization framework that easily allows to develop many extensions…
- … and that also allows to combine these extensions.

A new version of the original RGCCA package (with sequential RGCCA/SGCCA) was, not so long ago, submitted to the CRAN.

https://github.com/rgcca-factory/RGCCA

The RGCCA framework is:

- General as in encompasses a large number of methods in the multi-block literature.
- Very flexible thanks to its optimization framework that easily allows to develop many extensions…
- … and that also allows to combine these extensions.

A new version of the original RGCCA package (with sequential RGCCA/SGCCA) was, not so long ago, submitted to the CRAN.

It came with a submission to the Journal of Statistical Software (Girka et al., submitted).

https://github.com/rgcca-factory/RGCCA

RGCCA

<u>The RGCCA framework is:</u>

- ❖ General as in encompasses a large number of methods in the multi-block literature.
- ❖ Very flexible thanks to its optimization framework that easily allows to develop many extensions…
- ❖ … and that also allows to combine these extensions.

A new version of the original RGCCA package (with sequential RGCCA/SGCCA) was, not so long ago, submitted to the CRAN.

It came with a submission to the Journal of Statistical Software (Girka et al., submitted).

For extensions mentioned in this presentation and that are in development, you can find them one of the many branches of the github repository:
https://github.com/rgcca-factory/RGCCA

RGCCA

<u>The RGCCA framework is:</u>

Arthur TENENHAUS
Laboratoire Des Signaux Et Systèmes, CentraleSupélec

Fabien GIRKA
Laboratoire Des Signaux Et Systèmes, CentraleSupélec

- ❖ General as in encompasses a large number of methods in the multi-block literature.
- ❖ Very flexible thanks to its optimization framework that easily allows to develop many extensions.
- ❖ … and that also allows to combine these extensions.

Laurent LE BRUSQUET
Laboratoire Des Signaux Et Systèmes, CentraleSupélec

A new version of the original RGCCA package (with sequential RGCCA/SGCCA) was, not so long ago, submitted to the CRAN.

Etienne Camenen
INSERM, Hôpital Saint-Louis AP-HP

It came with a submission to the Journal of Statistical Software (Girka et al., submitted).

For extensions mentioned in this presentation and that are in development, you can find them one of the many branches of the github repository:
https://github.com/rgcca-factory/RGCCA

Caroline PELTIER
CNRS, INRAE

Here are the contributors of the actual version of the package !!

Vincent GUILLEMOT
Institut Pasteur, Bioinformatics and Biostatistics Hub

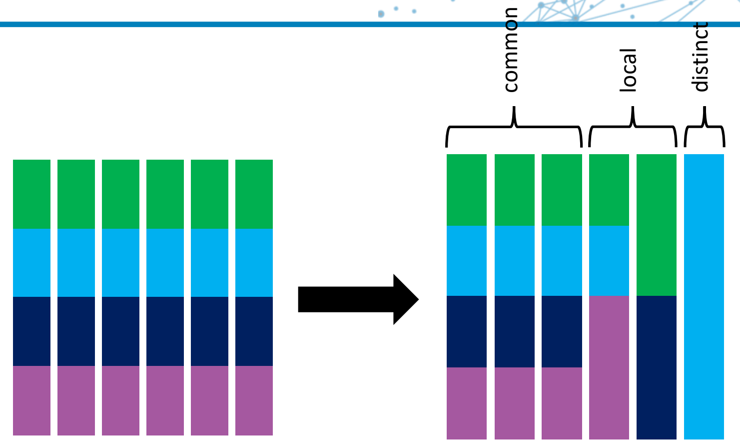
Arnaud GLOAGUEN
Centre National de Recherche en Génomique Humaine, CEA

INSTITUT FRANÇAIS DE BIOINFORMATIQUE
elixir FRANCE

Other perpectives ?

Other perpectives ?

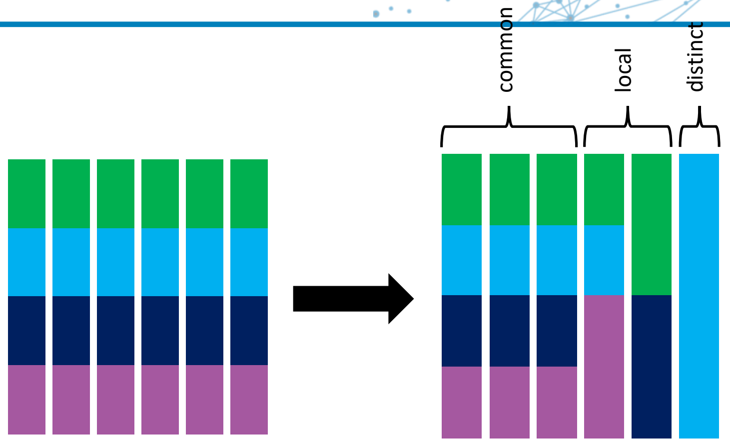
Axe 1: Use comon and specific information



Other perpectives ?

<u>Axe 1: Use comon and specific information</u>

Vary the combination of omics data from which components are built.

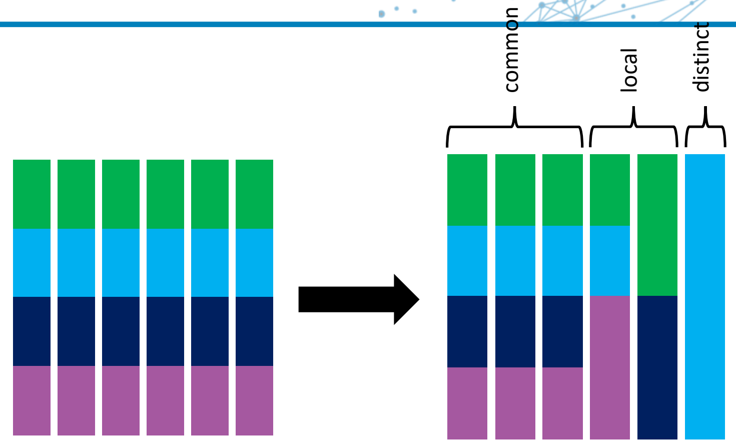


Other perpectives ?

Axe 1: Use comon and specific information

Vary the combination of omics data from which components are built.

Improve the interpretaion of components.

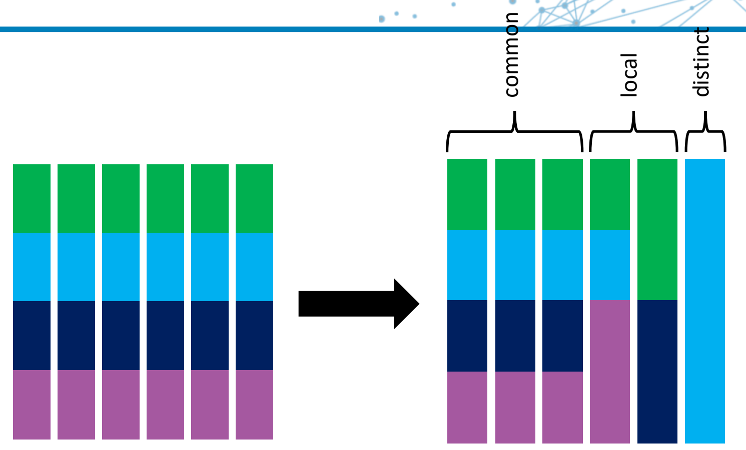


Other perpectives ?

Axe 1: Use comon and specific information

Vary the combination of omics data from which components are built.

Improve the interpretaion of components.



Axe 2: Include the appartenance of each variable to a biological pathway.

Divide each omic matrix by biological pathways.

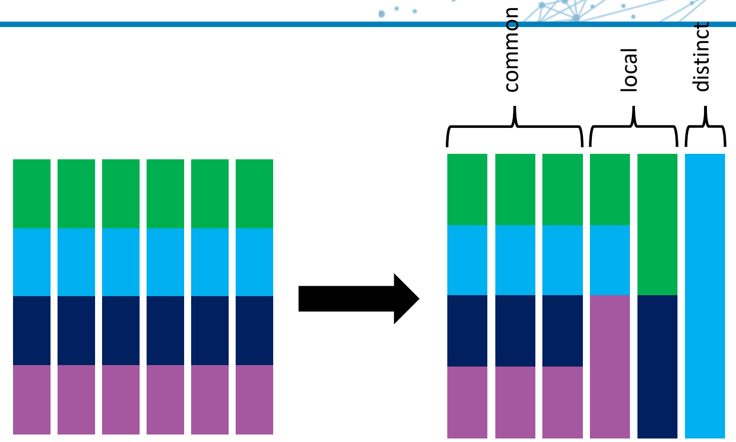
Allow to identify most important pathways (With L1 norm).

Other perspectives ?

Axe 1: Use comon and specific information

Vary the combination of omics data from which components are built.

Improve the interpretaion of components.



Axe 2: Include the appartenance of each variable to a biological pathway.

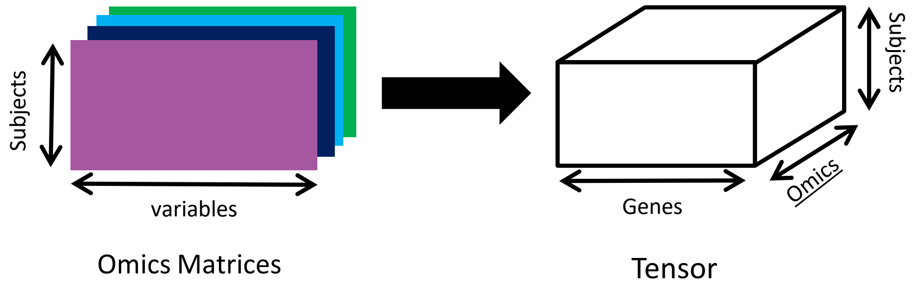
Divide each omic matrix by biological pathways.

Allow to identify most important pathways (With L1 norm).

Axe 3: Link variables across different omics

Regroup omic matrices along the third dimension (ex: by genes) to create a tensor.

Permet d'ajouter une notion biologique dans la définition du modèle



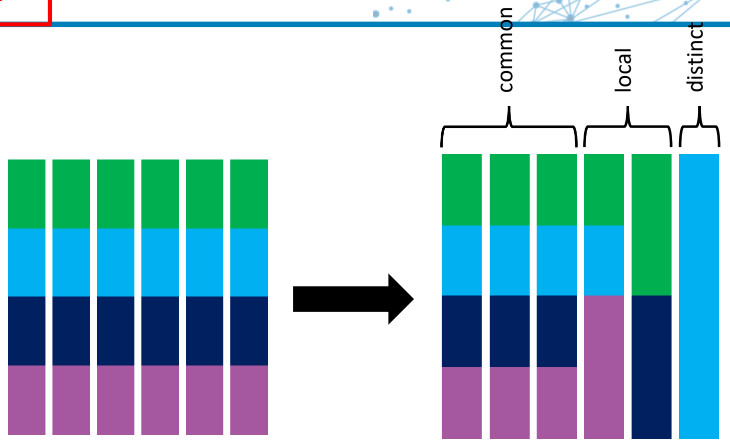
Courtesy to Vincent Le Goff.

Other perspectives ?

PhD of Vincent LE GOFF

Supervised by:

- **Edith Le Floch**
- **Vincent Guillemot**
- **Arnaud Gloaguen**



Axe 1: Use comon and specific information

Vary the combination of omics data from which components are built.

Improve the interpretaion of components.

Axe 2: Include the appartenance of each variable to a biological pathway.

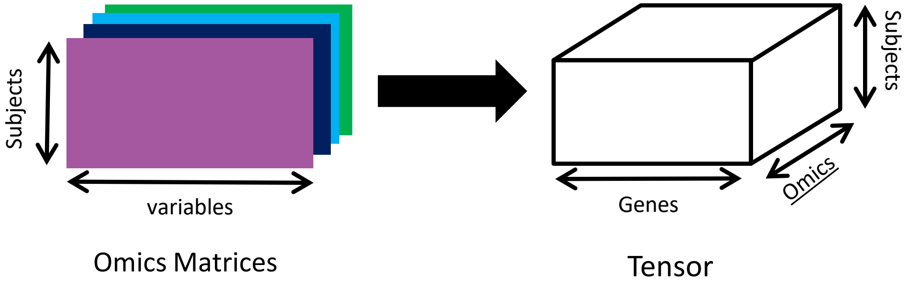
Divide each omic matrix by biological pathways.

Allow to identify most important pathways (With L1 norm).

Axe 3: Link variables across different omics

Regroup omic matrices along the third dimension (ex: by genes) to create a tensor.

Permet d'ajouter une notion biologique dans la définition du modèle



Courtesy to Vincent Le Goff.

Acknowledgment

Data (CT-ScarBip Project)

Andrée Delahaye-Duriez

Pierre-Eric Lutz

Amazigh Mokhtari

David Cohen

Bruno Etain

Cynthia Marie-Clair

El Chérif Ibrahim

Belzeaux Raoul

Gascon Gonzalo

Isabelle Mansuy

Summer School SIB/IFB:

Florence Mehl

Olivier Sand

Grégoire Rossier

Jimmy Vandel

Guillemette Marot

Marie-Galadriel Briere

Lucie Khamvongsa-Charbonnier

Morgane Terezol

Anaïs Baudot

Maxime Delmas

Jean-Clément Gallardo

Marco Pagni

Statomique:

Guillemette Marot (again !)

Christelle Hennequet-Antier

Julie Aubert

Marie-Agnès Dillies

Helene Touzet

Justine Merlan

Slides:

Arthur Tenenhaus

Laurent Le Brusquet

Julien Bect

Vincent Le Goff

Edith Le Floch

ETBII:

Lucie Khamvongsa-Charbonnier

Hélène Chiapello

Olivier Sand

Jimmy Vandel

Vincent Guillemot

Marie-Galadriel Briere

Morgane Terezol

Alban Gaignard

Olivier Dameron

Pierre Larmande

INSTITUT FRANÇAIS DE BIOINFORMATIQUE

References

Chegraoui, H. et al. (2023) 'Integrating multiomics and prior knowledge: a study of the Graphnet penalty impact', Bioinformatics, 39(8), p. btad454. Available at: https://doi.org/10.1093/bioinformatics/btad454.

Girka, F. (2023) Développement de méthodes statistiques/ML pour l'identification de biomarqueurs longitudinaux multimodaux. Application à la compréhension des mécanismes d'évolution de la sclérose en plaques. These en préparation. université Paris-Saclay. Available at: https://www.theses.fr/s251361 (Accessed: 20 November 2023).

Girka, F. et al. (2024) 'Tensor generalized canonical correlation analysis', Information Fusion, 102, p. 102045. Available at: https://doi.org/10.1016/j.inffus.2023.102045.

Girka, F. et al. (submitted to JSS) 'Multiblock data analysis with the RGCCA package', Journal of Statistical Software [Preprint].

Gloaguen, A. (2020) A statistical and computational framework for multiblock and multiway data analysis. phdthesis. Université Paris-Saclay. Available at: https://theses.hal.science/tel-03044035 (Accessed: 20 November 2023).

Gloaguen, A. et al. (2022) 'Multiway generalized canonical correlation analysis', Biostatistics (Oxford, England), 23(1), pp. 240–256. Available at: https://doi.org/10.1093/biostatistics/kxaa010.

Goujon, E. (In progress) Identification de marqueurs moléculaires de la radiation à faibles doses à partir de données multi-omiques. These en préparation. université Paris-Saclay. Available at: https://www.theses.fr/s357653 (Accessed: 20 November 2023).

Guigui, N. et al. (2019) 'Network Regularization in Imaging Genetics Improves Prediction Performances and Model Interpretability on Alzheimer's Disease', in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI), Venice, Italy: IEEE, pp. 1403–1406. Available at: https://doi.org/10.1109/ISBI.2019.8759593.

Guillemot, V. et al. (2021) 'Introducing group-sparsity and orthogonality constraints in RGCCA'.

Löfstedt, T. et al. (2016) 'A general multiblock method for structured variable selection'. arXiv. Available at: http://arxiv.org/abs/1610.09490 (Accessed: 10 April 2023).

Peltier, C. et al. (2023) 'Missing Values in RGCCA: Algorithms and Comparisons', in L. Radomir et al. (eds) State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM): Methodological Extensions and Applications in the Social Sciences and Beyond. Cham: Springer International Publishing (Springer Proceedings in Business and Economics), pp. 9–14. Available at: https://doi.org/10.1007/978-3-031-34589-0_2.

Sort, L., Brusquet, L.L. and Tenenhaus, A. (2023) 'Functional Generalized Canonical Correlation Analysis for studying multiple longitudinal variables'. arXiv. Available at: https://doi.org/10.48550/arXiv.2310.07330.

Tenenhaus, A. and Tenenhaus, M. (2011) 'Regularized Generalized Canonical Correlation Analysis', Psychometrika, 76(2), pp. 257–284. Available at: https://doi.org/10.1007/s11336-011-9206-8.

Tenenhaus, A. and Tenenhaus, M. (2014) 'Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis', European Journal of Operational Research, 238(2), pp. 391–403. Available at: https://doi.org/10.1016/j.ejor.2014.01.008.

Tenenhaus, A. et al. (2014) 'Variable selection for generalized canonical correlation analysis', Biostatistics, 15(3), pp. 569–583. Available at: https://doi.org/10.1093/biostatistics/kxu001.

Tenenhaus, A., Philippe, C. and Frouin, V. (2015) 'Kernel Generalized Canonical Correlation Analysis', Computational Statistics & Data Analysis, 90, pp. 114–131. Available at: https://doi.org/10.1016/j.csda.2015.04.004.

Tenenhaus, M., Tenenhaus, A. and Groenen, P.J.F. (2017) 'Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods', Psychometrika, 82(3), pp. 737–777. Available at: https://doi.org/10.1007/s11336-017-9573-x.

Thank you for your attention !