

Regularized Generalized Canonical Correlation Analysis (RGCCA), a flexible framework for multiblock (and multiway) data analysis: case study on a cohort of Major Depressive Disorder (MDD) patients and controls

Arnaud Gloaguen, Jimmy Vandel



Groupement
de recherche
BIM Bioinformatique moléculaire



Swiss Institute of
Bioinformatics





1. Introduction of the case study

Case Study: Major Depressive Disorder (MDD)



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. →

		BMI	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

> 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			

Low (≤ 25), High (≥ 30).

Relative to position on the DNAm chip.

> summary(DNAm_covariates_explored_female)

Sample_Group	BMI
control:50	Min. :16.37
mdd :37	1st Qu.:21.42
	Median :23.23
	Mean :23.83
	3rd Qu.:25.24
	Max. :39.54

BMI.bin
low :64
medium:19
high : 4

Age
Min. :21.00
1st Qu.:32.00
Median :45.00
Mean :43.52
3rd Qu.:53.50
Max. :71.00

Age.bin
<20 : 0
20-30:21
30-40:11
40-50:20
50-60:26
60-70: 8
>70 : 1

Age_bin

2:21

3:11

4:20

5:26

6: 8

7: 1

R04C01 :19

R05C01 :18

R06C01 :12

R07C01 :12

R03C01 :10

R02C01 : 8

(Other): 8

(Other)

:73

GR

(Other)

:73

CD4

CD8

MO

B

NK

GR

Min. :0.08709

Min. :0.02919

Min. :0.04979

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

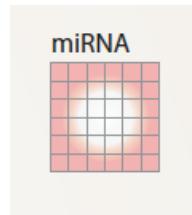
Relative to blood cell composition (T cells subsets, monocytes, B cells, NK cells and granulocytes) inferred from DNAm.



Case Study: Pre-processing

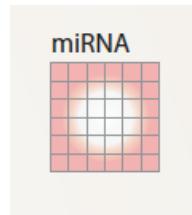


Case Study: Pre-processing





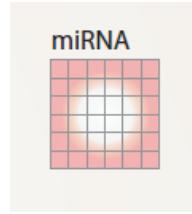
Case Study: Pre-processing



1. Remove miRNA with Nas.



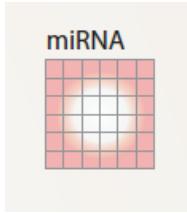
Case Study: Pre-processing



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



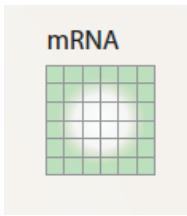
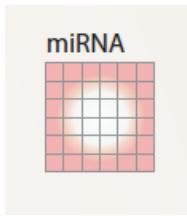
Case Study: Pre-processing



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).



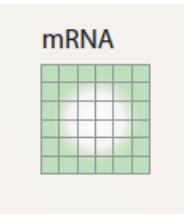
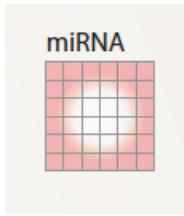
Case Study: Pre-processing



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).



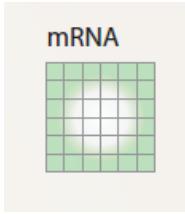
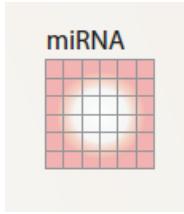
Case Study: Pre-processing



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).



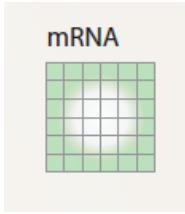
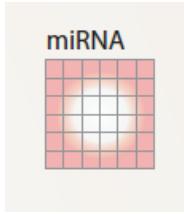
Case Study: Pre-processing



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

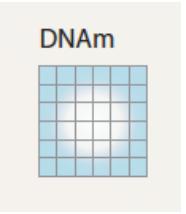
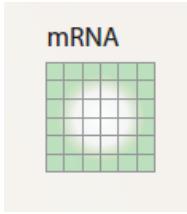
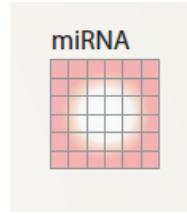


1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

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



Case Study: Pre-processing

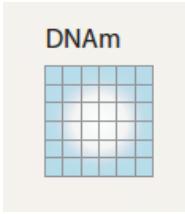
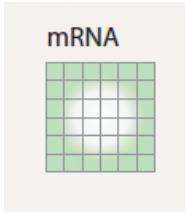
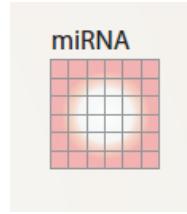


1. Remove miRNA with Nas.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

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



Case Study: Pre-processing

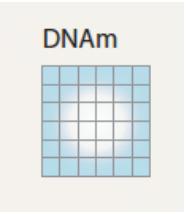
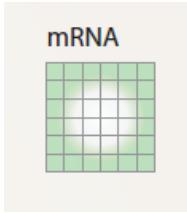
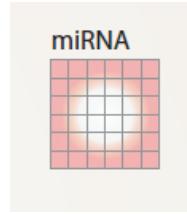


- 1. Remove miRNA with Nas.
- 2. Log Counts Per Million (logCPM) normalization.
- 3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).
- 1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
- 2. Keep only genes both present in males and females.
- 3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).
- 1. Normalization with Beta-Mixture Quantile (BMRQ) Normalization method (package ChAMP).

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



Case Study: Pre-processing

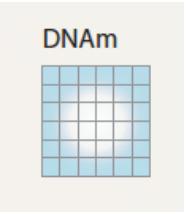
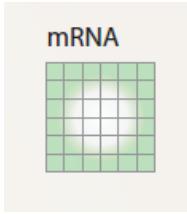
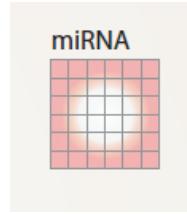


1. Remove miRNA with NAs.
2. Log Counts Per Million (logCPM) normalization.
3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
2. Keep only genes both present in males and females.
3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).
1. Normalization with Beta-Mixture Quantile (BMRQ) Normalization method (package ChAMP).
2. Remove duplicated samples.

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



Case Study: Pre-processing

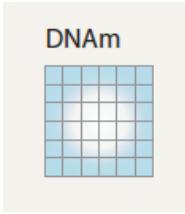
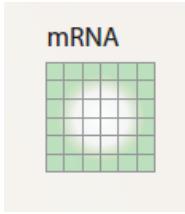
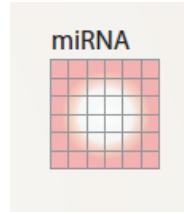


1. Remove miRNA with NAs.
 2. Log Counts Per Million (logCPM) normalization.
 3. Remove miRNA with a least one count below 0 (in the end 350 variables remain).
1. Normalization with Variance Stabilizing Transformations (VST; package DESeq2).
 2. Keep only genes both present in males and females.
 3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).
1. Normalization with Beta-Mixture Quantile (BMRQ) Normalization method (package ChAMP).
 2. Remove duplicated samples.
 3. Keep 2000 most variable genes according to the Median Absolute Deviation (MAD).

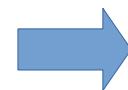
$MAD = \text{median}(|x_i - \text{median}(\mathbf{x})|)$, it is a robust estimation of the standard deviation.



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 (BMRQ) 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.

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



2. Unsupervised analysis with one-block



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, \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.



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.

{



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.

$\left\{ H_0: (x_1, \dots, x_n) \text{ and } (y_1, \dots, y_m) \text{ comes from the same distribution.} \right.$



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, \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.

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

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.

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

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 $\frac{nm+1}{2}$.



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.

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

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, \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.

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

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:

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



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.

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

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:

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

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



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.

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

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:

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

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



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.

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

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:

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

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



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.

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

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:

$$\begin{aligned} R_1 &= 1 + 2 + \dots + n = \frac{n(n+1)}{2} & \rightarrow U_1 &= nm + \frac{n(n+1)}{2} - R_1 = nm \\ R_2 &= (n+1) + (n+2) + \dots + (n+m) = \frac{m((n+1) + (n+m))}{2} & \rightarrow U_2 &= nm + \frac{m(m+1)}{2} - R_2 = 0 \end{aligned} \quad \left. \begin{array}{l} \\ \end{array} \right\}$$



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.

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

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:

$$\begin{aligned} R_1 &= 1 + 2 + \dots + n = \frac{n(n+1)}{2} & \rightarrow U_1 &= nm + \frac{n(n+1)}{2} - R_1 = nm \\ R_2 &= (n+1) + (n+2) + \dots + (n+m) = \frac{m((n+1) + (n+m))}{2} & \rightarrow U_2 &= nm + \frac{m(m+1)}{2} - R_2 = 0 \end{aligned} \quad \left. \begin{array}{l} \rightarrow U_1 = nm + \frac{n(n+1)}{2} - R_1 = nm \\ \rightarrow U_2 = nm + \frac{m(m+1)}{2} - R_2 = 0 \end{array} \right\} \rightarrow U = 0$$



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.

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

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:

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

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

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

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

The test is likely to be rejected.



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: $\left\{ \begin{array}{l} H_0: \beta_1 = 0, \\ \end{array} \right.$



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:
$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$



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:



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:

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



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:

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

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



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:

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

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:

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

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 $\rightarrow RSS_0 \gg RSS_1$



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:
$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

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 $\rightarrow RSS_0 \gg RSS_1 \rightarrow F \sim \frac{RSS_0}{RSS_1} \times (n - 2) \gg n - 2$



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:

$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

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 $\rightarrow RSS_0 \gg RSS_1 \rightarrow F \sim \frac{RSS_0}{RSS_1} \times (n - 2) \gg n - 2 \rightarrow$ 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:

$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

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 $\rightarrow RSS_0 \gg RSS_1 \rightarrow F \sim \frac{RSS_0}{RSS_1} \times (n - 2) \gg n - 2 \rightarrow$ 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:

$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

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 $\rightarrow RSS_0 \gg RSS_1 \rightarrow F \sim \frac{RSS_0}{RSS_1} \times (n - 2) \gg n - 2 \rightarrow$ The test is likely to be rejected.

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



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:

$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

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 $\rightarrow RSS_0 \gg RSS_1 \rightarrow F \sim \frac{RSS_0}{RSS_1} \times (n - 2) \gg n - 2 \rightarrow$ The test is likely to be rejected.

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



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:

$$\begin{cases} H_0: \beta_1 = 0, \\ H_1: \beta_1 \neq 0 \end{cases}$$

With the statistic:

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

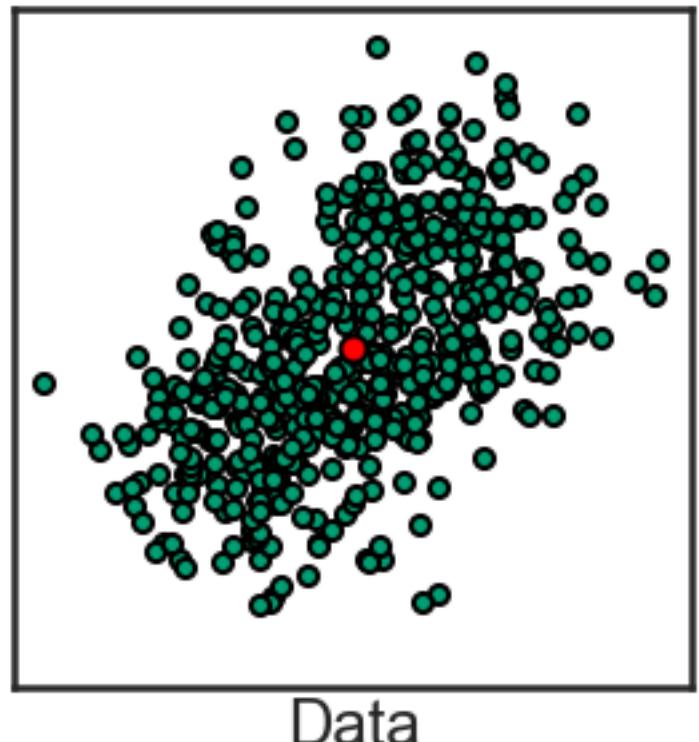
It is possible to show that F follows an F-distribution of 1 and $n - 2$ degrees of freedom.

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

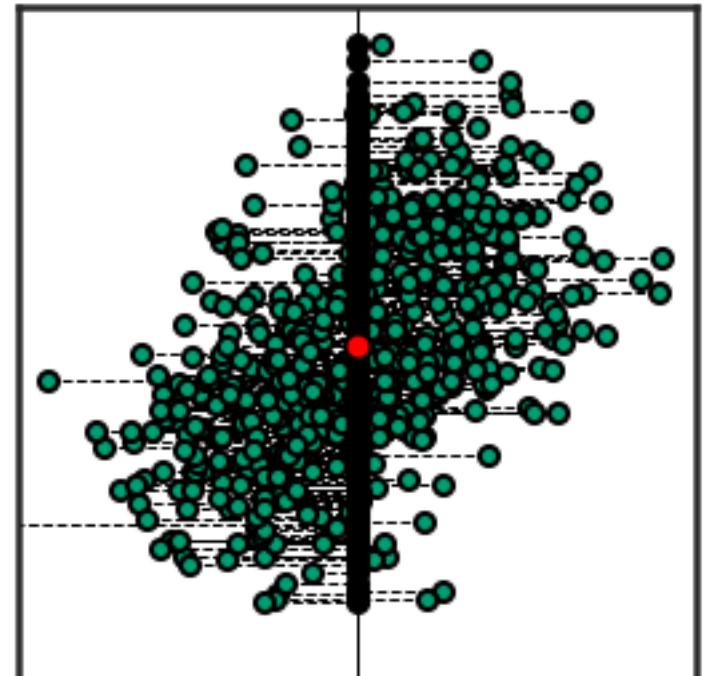
If the variables are NOT strongly linked $\rightarrow RSS_0 \sim RSS_1 \rightarrow F \sim 0 \rightarrow$ The test is likely to be accepted.



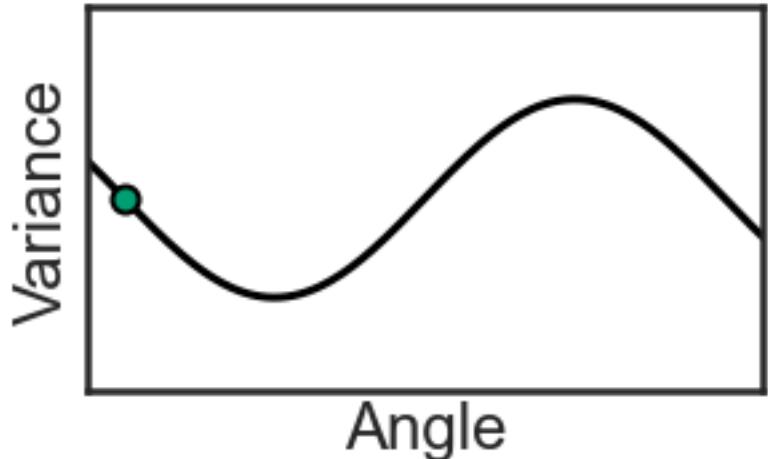
Principle of Principal Component Analysis (PCA)



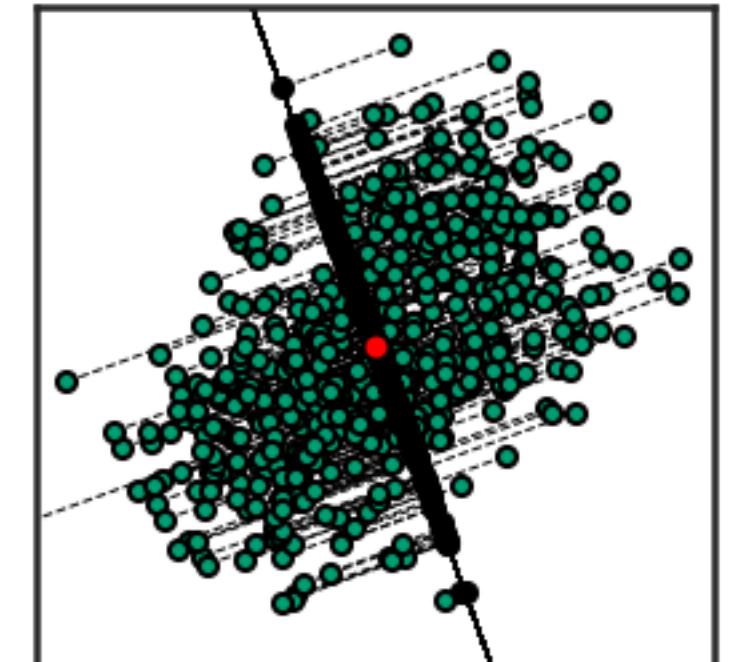
Principle of Principal Component Analysis (PCA)



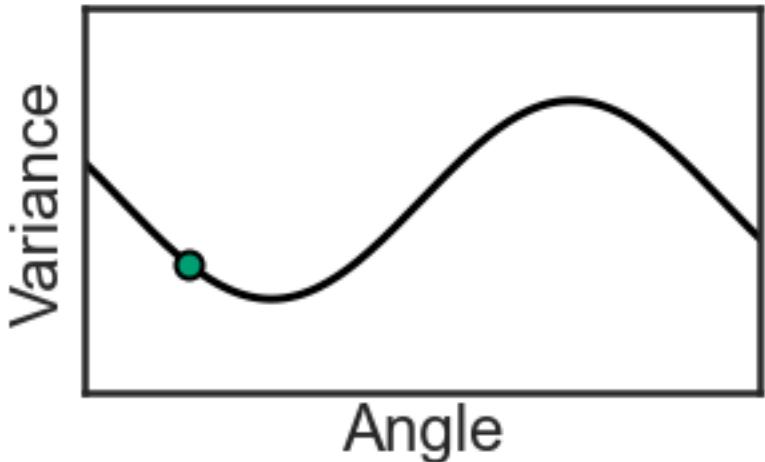
Data, mean and projection



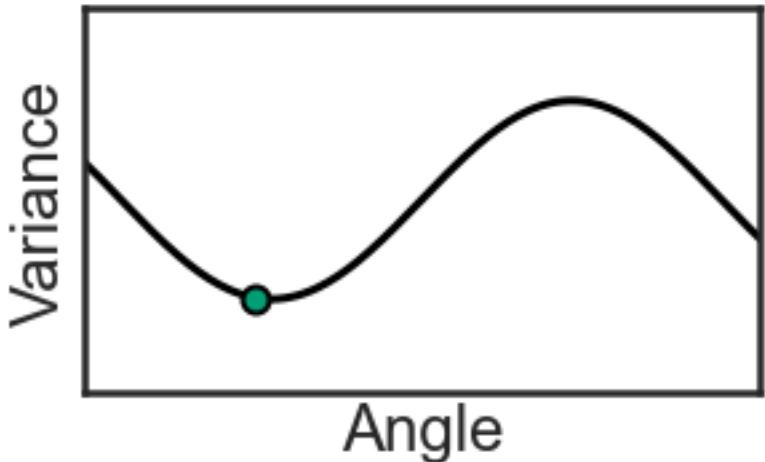
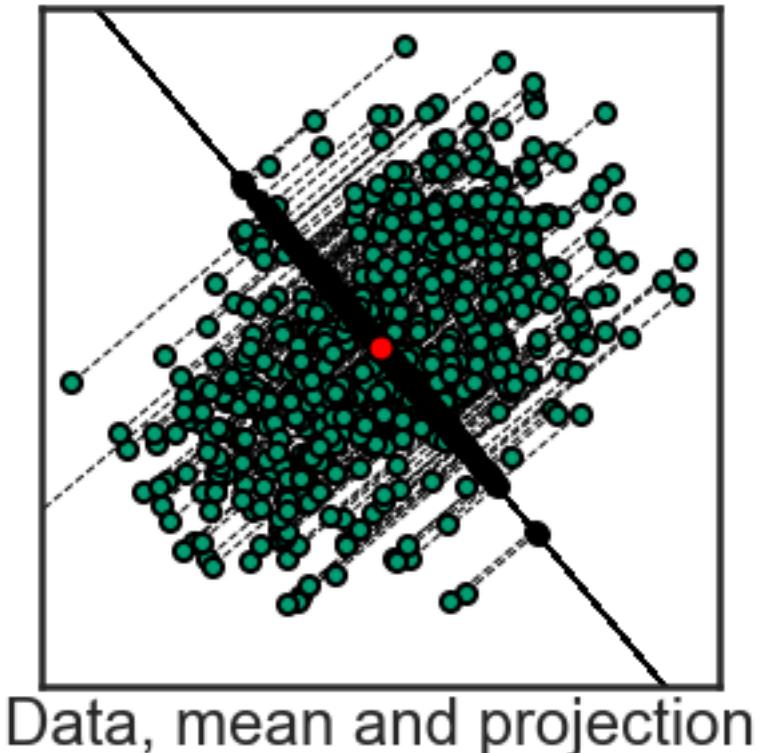
Principle of Principal Component Analysis (PCA)



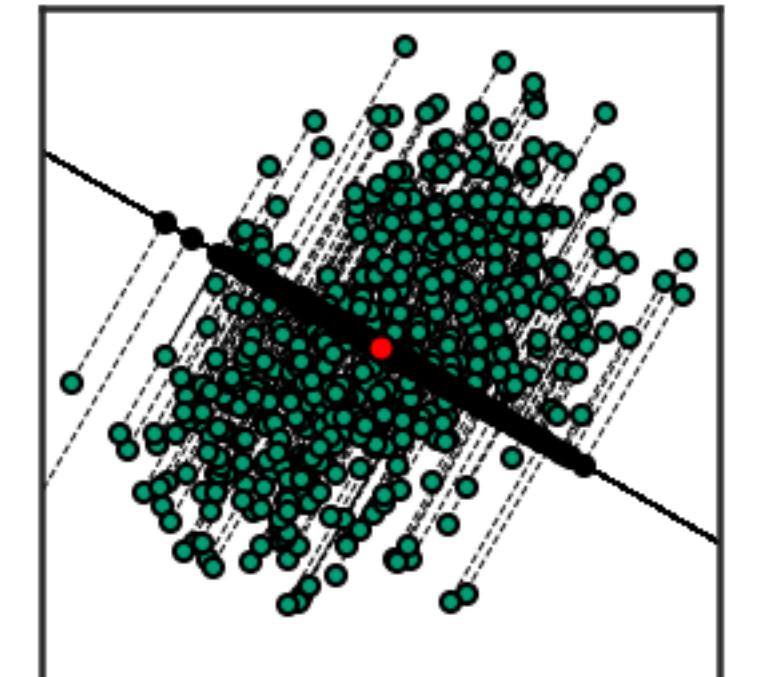
Data, mean and projection



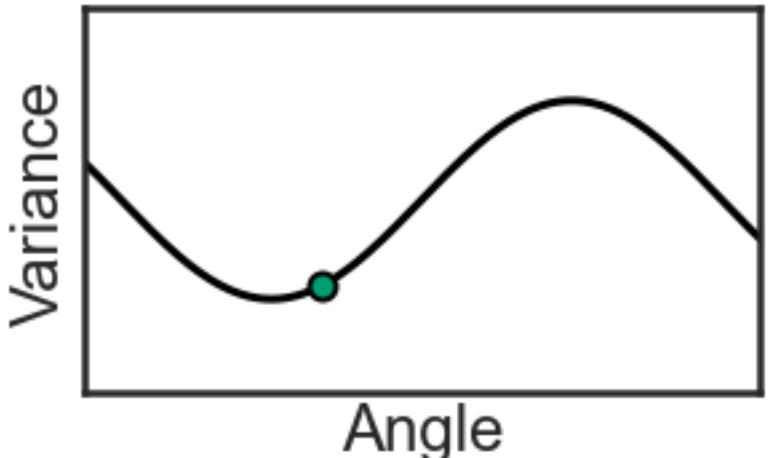
Principle of Principal Component Analysis (PCA)



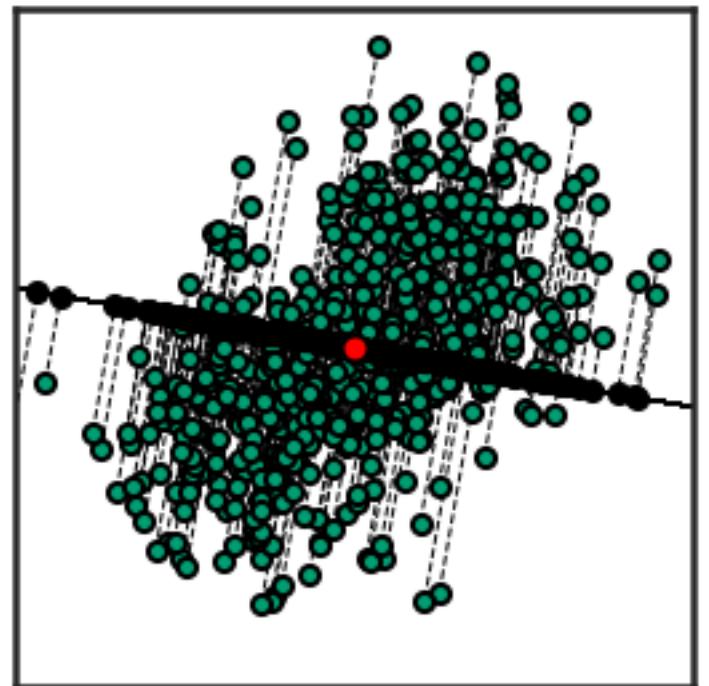
Principle of Principal Component Analysis (PCA)



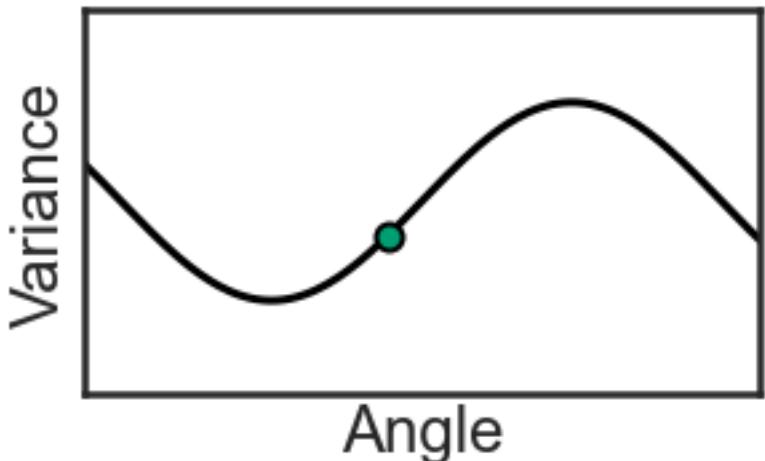
Data, mean and projection



Principle of Principal Component Analysis (PCA)

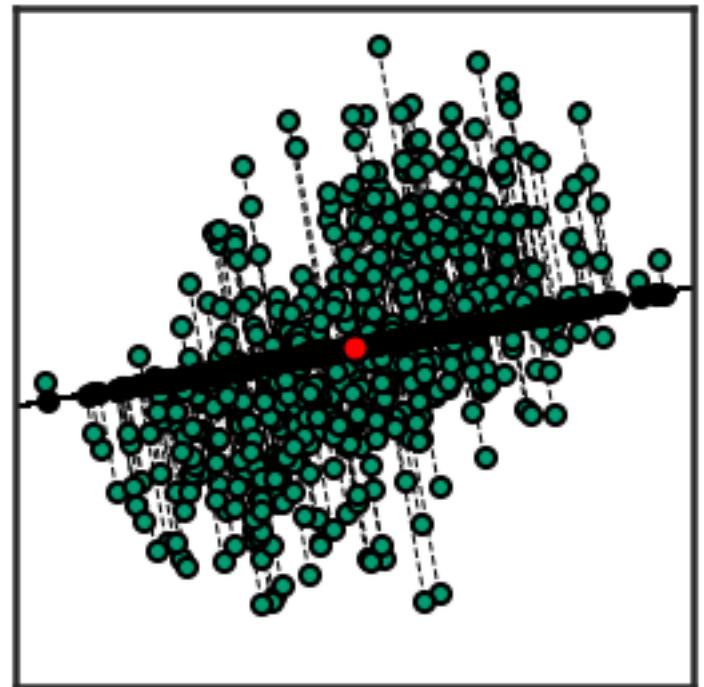


Data, mean and projection

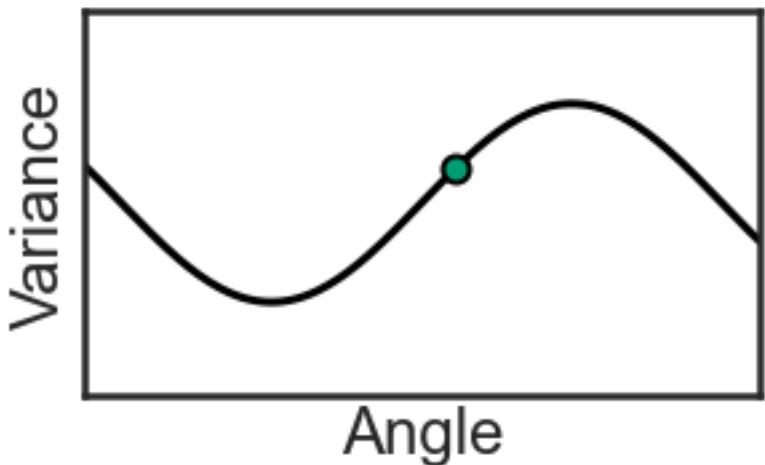




Principle of Principal Component Analysis (PCA)

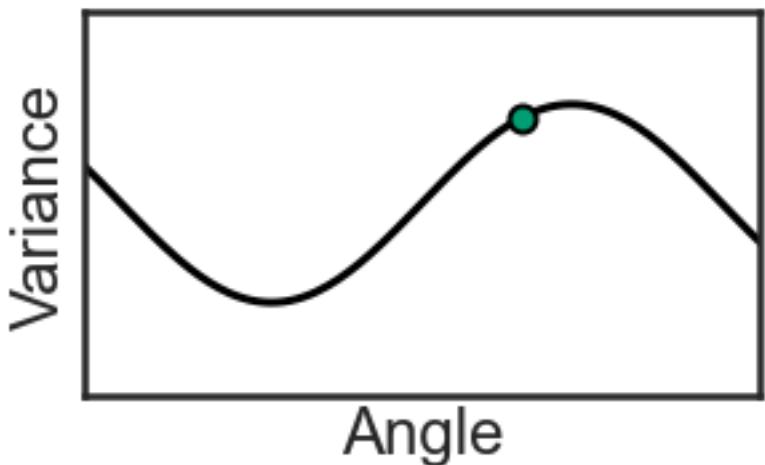
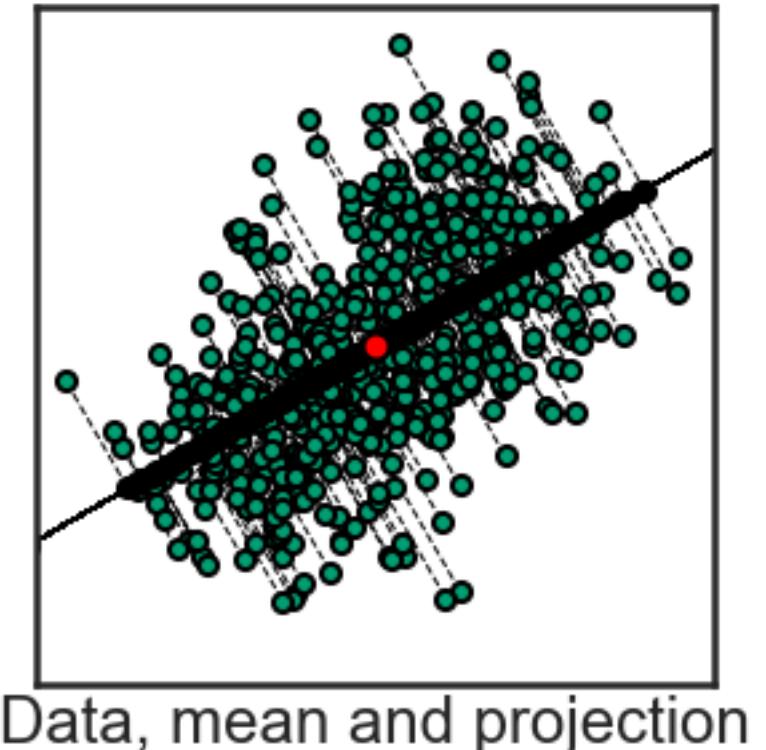


Data, mean and projection

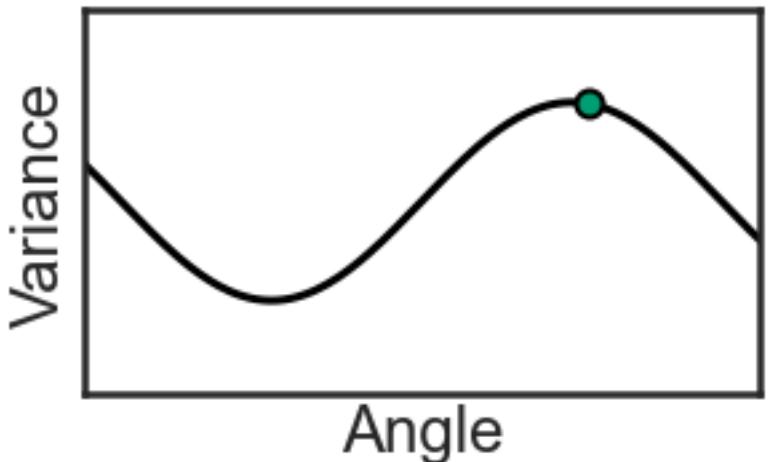
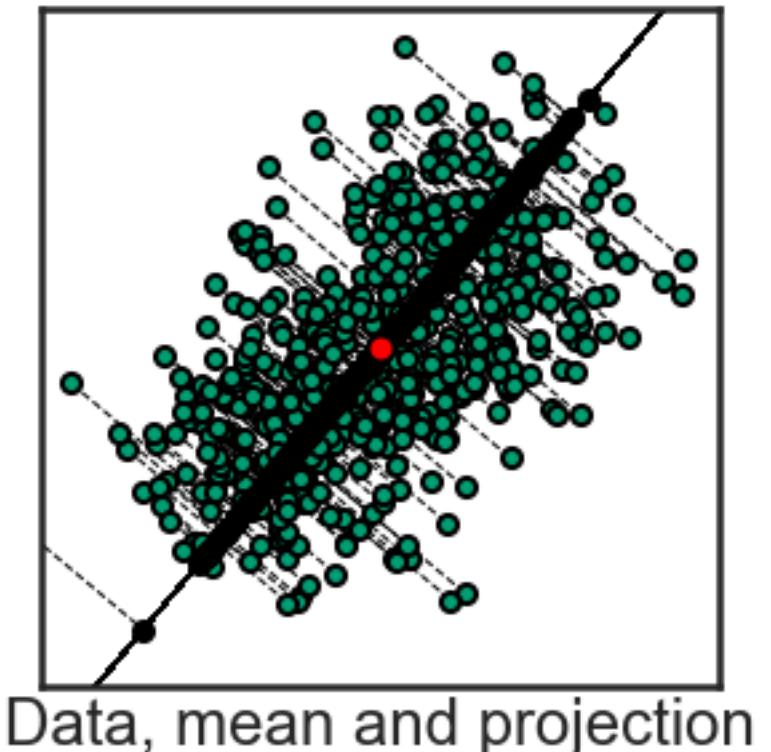




Principle of Principal Component Analysis (PCA)

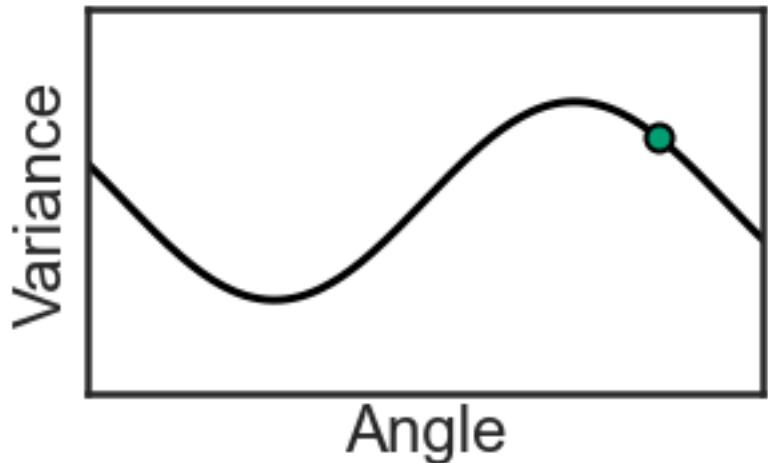
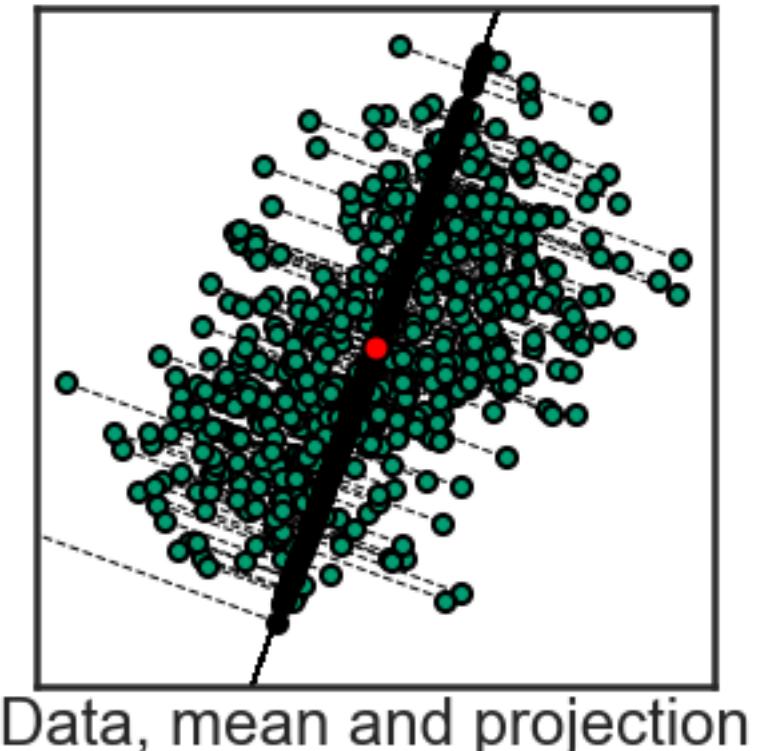


Principle of Principal Component Analysis (PCA)

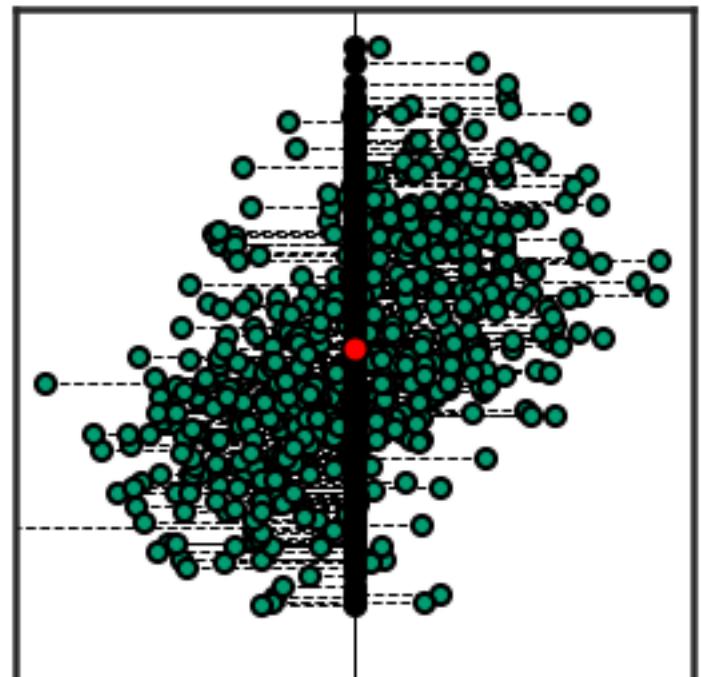




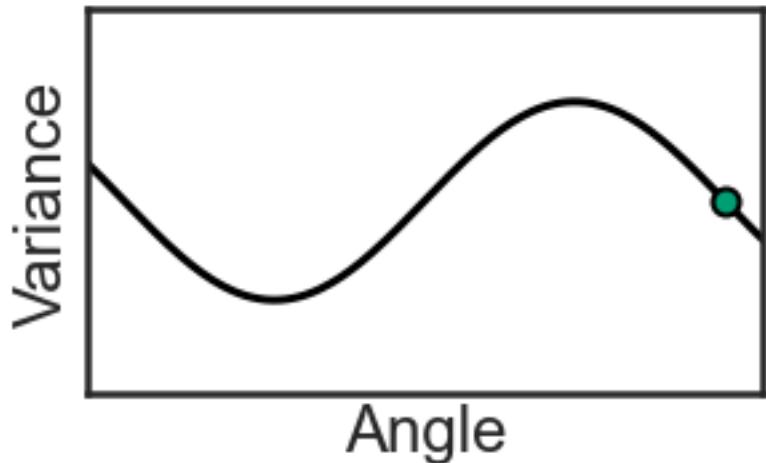
Principle of Principal Component Analysis (PCA)



Principle of Principal Component Analysis (PCA)

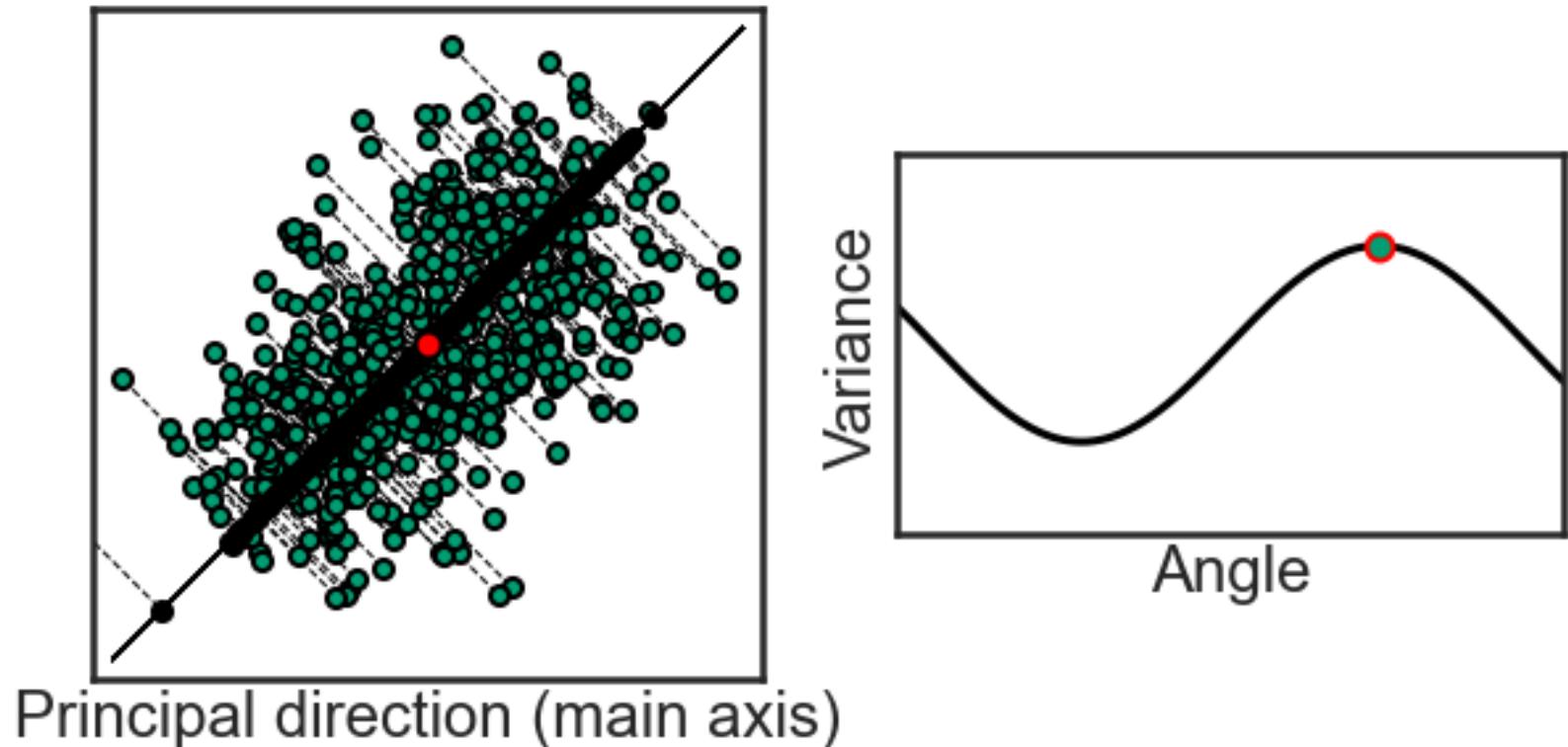


Data, mean and projection



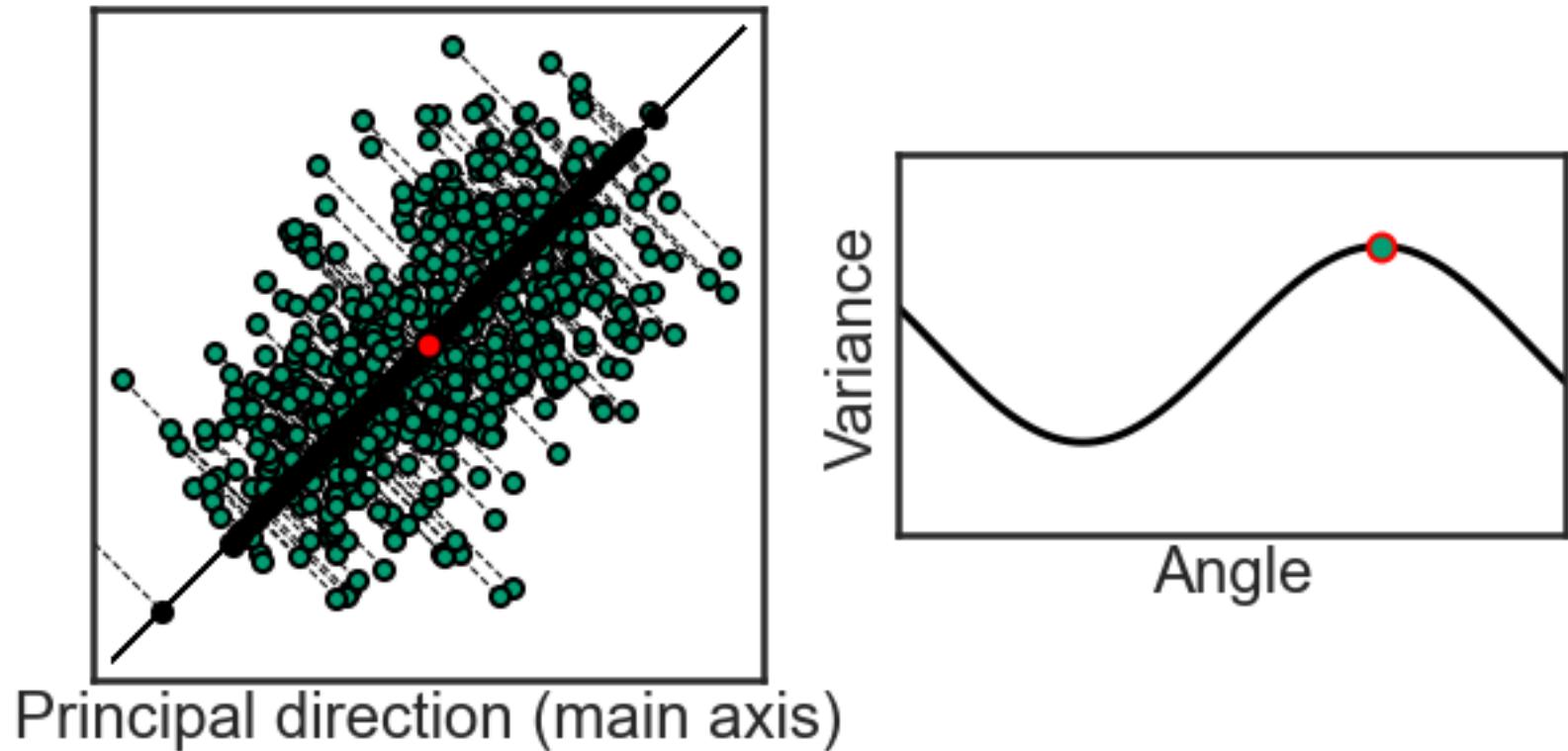


Principle of Principal Component Analysis (PCA)





Principle of Principal Component Analysis (PCA)



How can we estimate this projection ?

Courtesy to Arthur Tenenhaus, Laurent Le Busquet and Julien Bect.



Total Variance



Total Variance

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$



Total Variance

$$\mathbf{X} = \begin{pmatrix} & & p \text{ variables} \\ \xleftarrow{x_{11} \quad \cdots \quad x_{1p}} & & \xrightarrow{\vdots \quad \ddots \quad \vdots} \\ & & \xleftarrow{x_{n1} \quad \cdots \quad x_{np}} \end{pmatrix}$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

p variables



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix}$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

Total Variance (TV)



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\text{Total Variance (TV)} = \sum_{j=1}^p \text{Variance}(\text{Variable}_j)$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\text{Total Variance (TV)} = \sum_{j=1}^p \text{Variance}(\text{Variable}_j) = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\text{Total Variance (TV)} = \sum_{j=1}^p \text{Variance}(\text{Variable}_j) = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\begin{aligned} \text{Total Variance (TV)} &= \sum_{j=1}^p \text{Variance}(\text{Variable}_j) = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\| \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \right\|_2^2 \end{aligned}$$



Total Variance

n individuals/subjects

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\begin{aligned} \text{Total Variance (TV)} &= \sum_{j=1}^p \text{Variance}(\text{Variable}_j) = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\| \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \right\|_2^2 = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \bar{\mathbf{x}}\|_2^2 \end{aligned}$$



Total Variance

n individuals/subjects

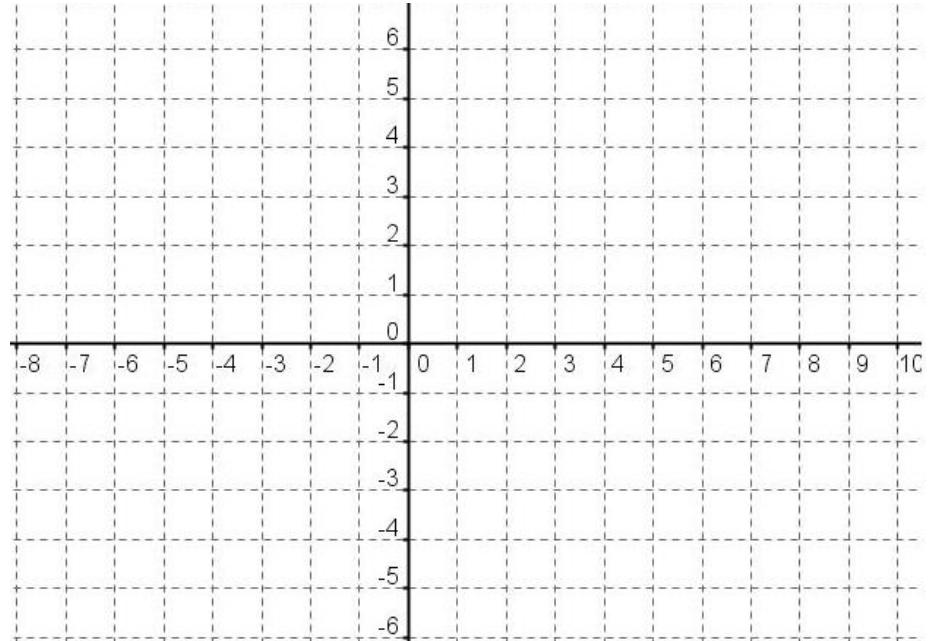
$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} \text{ where } \mathbf{x}_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

$$\begin{aligned} \text{Total Variance (TV)} &= \sum_{j=1}^p \text{Variance}(\text{Variable}_j) = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - \bar{x}_j)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left\| \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} - \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix} \right\|_2^2 = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i - \bar{\mathbf{x}}\|_2^2 \end{aligned}$$

Here, we suppose that every variable is centered $\Rightarrow \text{TV} = \frac{1}{n} \sum_{i=1}^n \|\mathbf{x}_i\|_2^2$

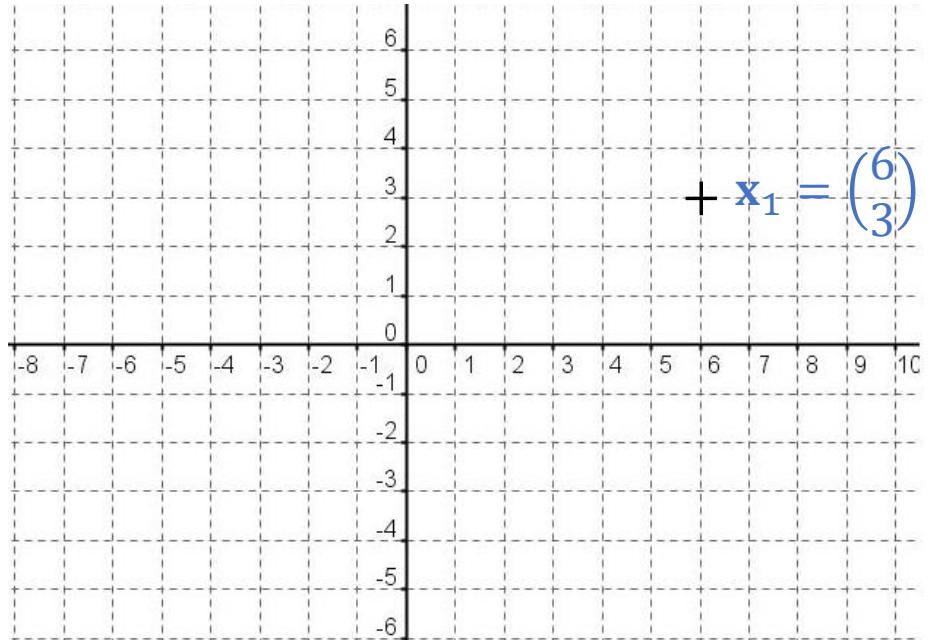


Total Variance – Example in 2D



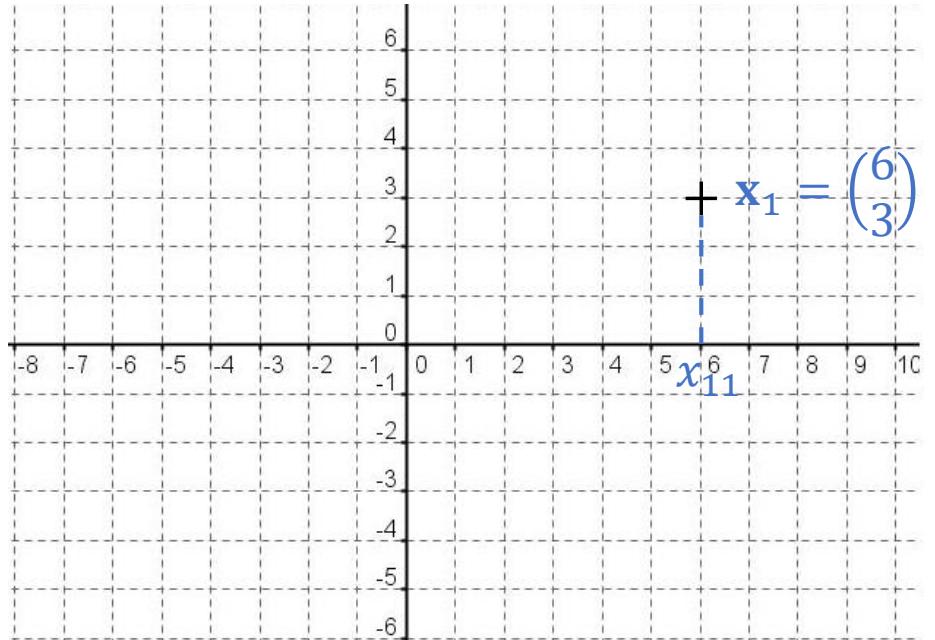


Total Variance – Example in 2D



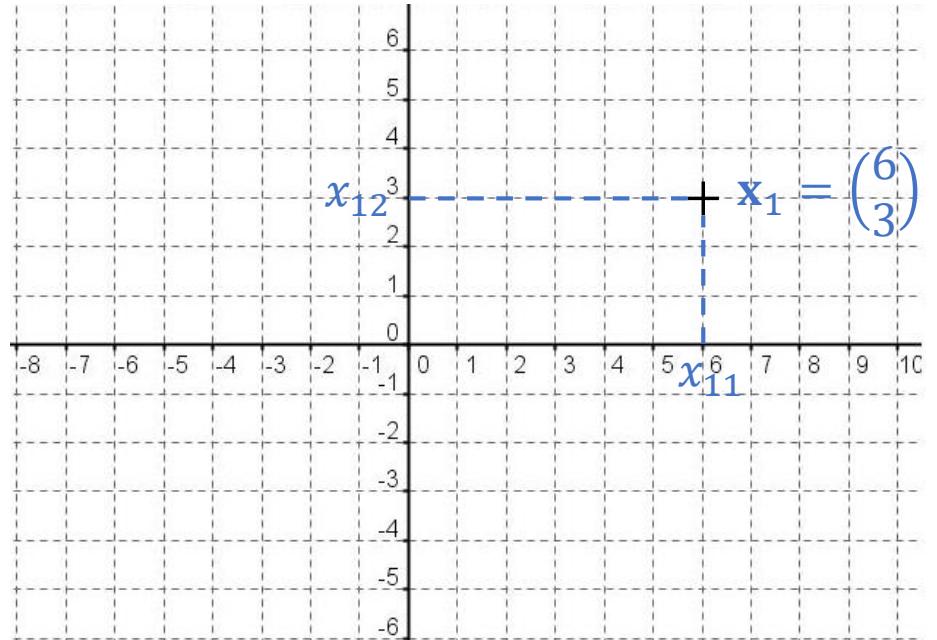


Total Variance – Example in 2D





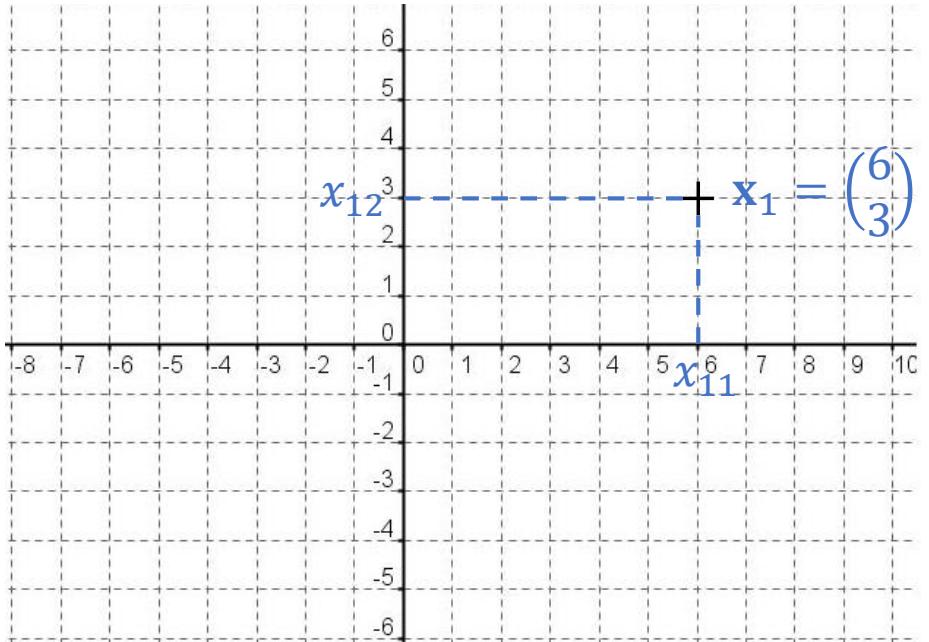
Total Variance – Example in 2D





Total Variance – Example in 2D

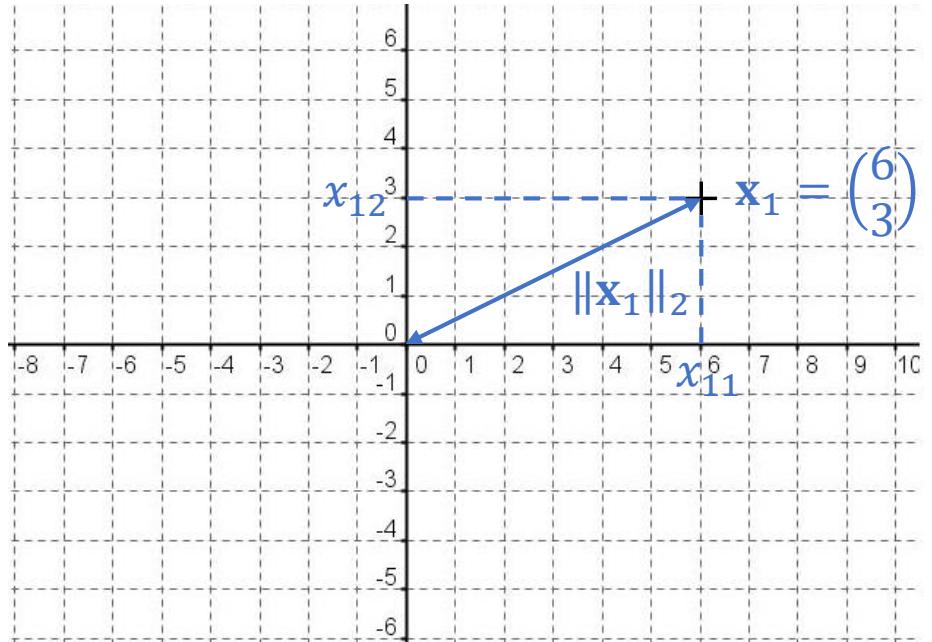
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$





Total Variance – Example in 2D

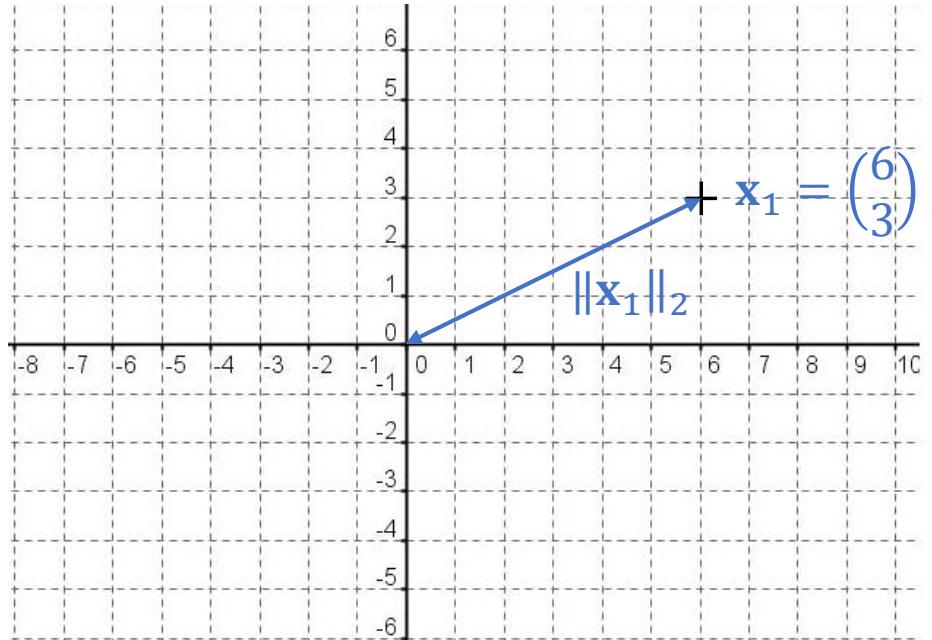
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$





Total Variance – Example in 2D

$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$

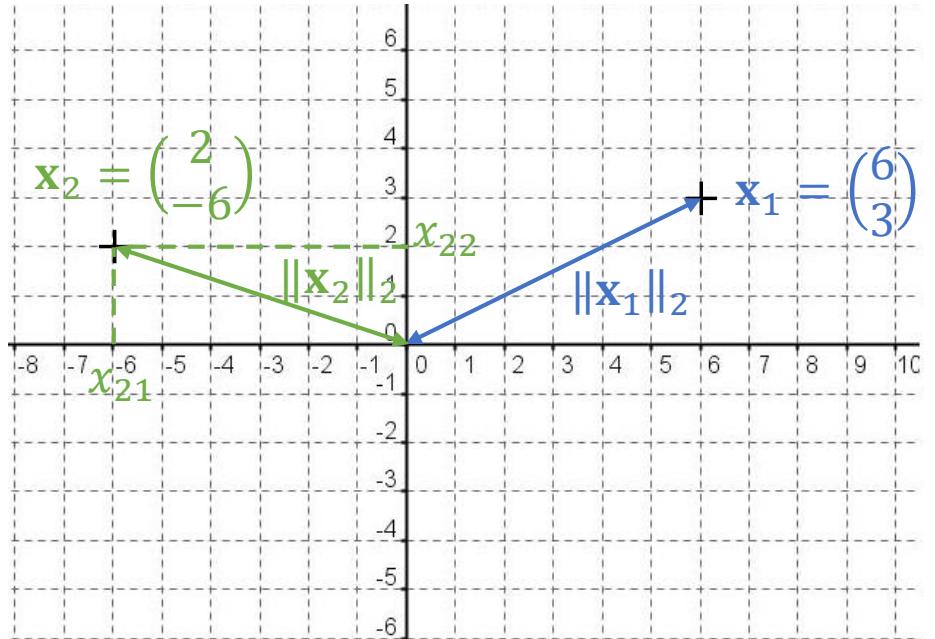




Total Variance – Example in 2D

$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$

$$\|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

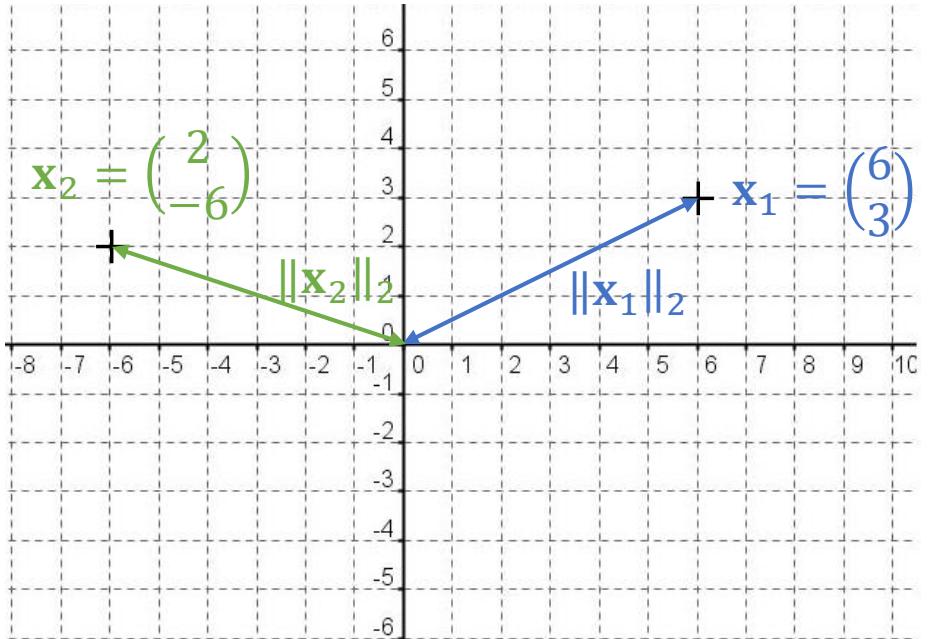




Total Variance – Example in 2D

$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$

$$\|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

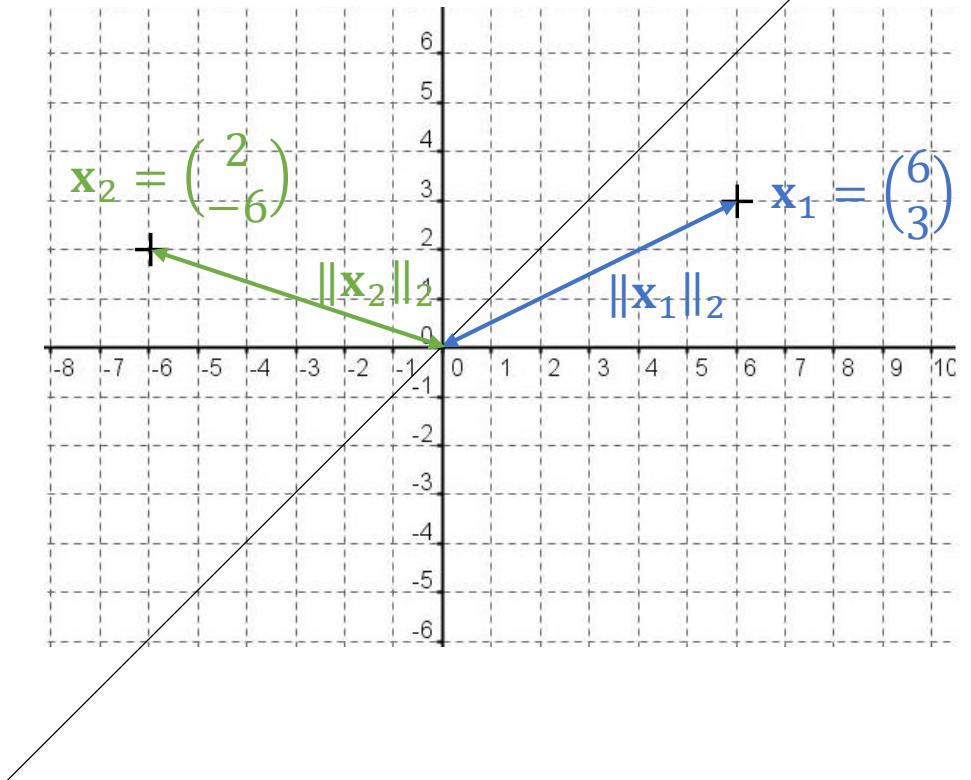




Total Variance – Example in 2D

$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$

$$\|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

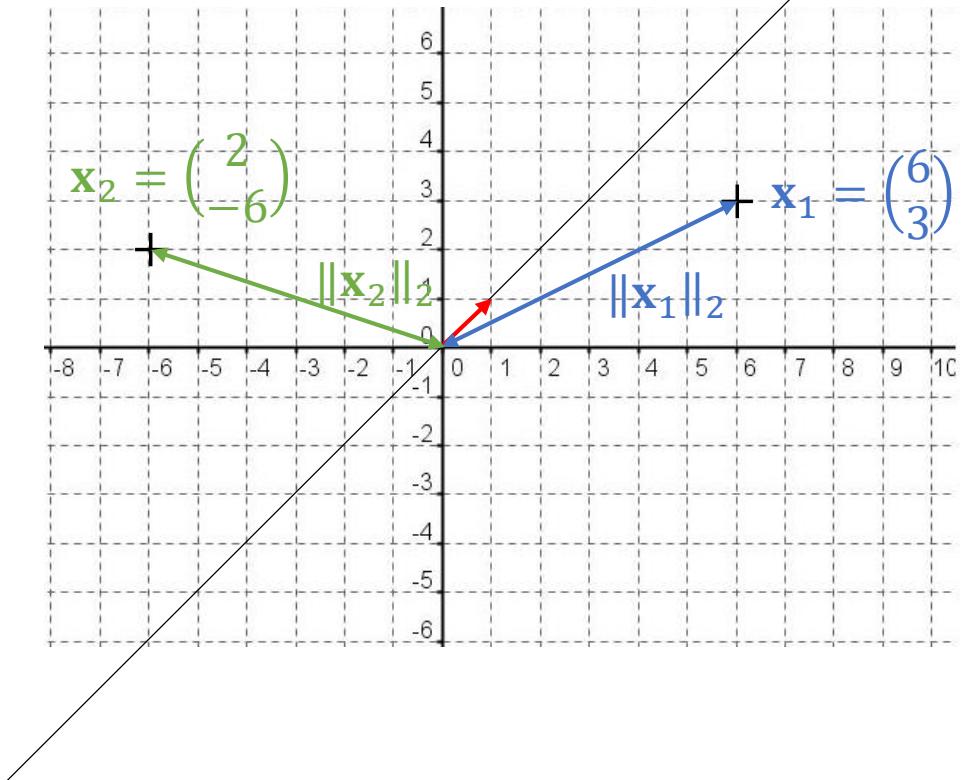




Total Variance – Example in 2D

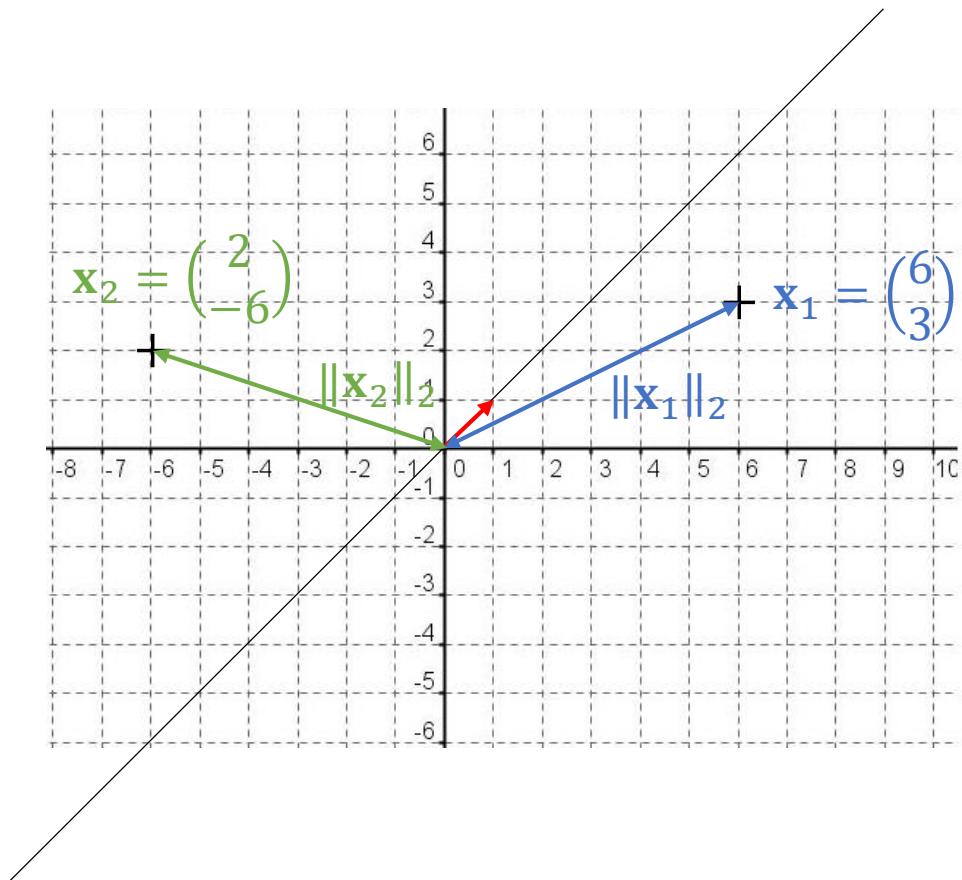
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2}$$

$$\|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$





Total Variance – Example in 2D

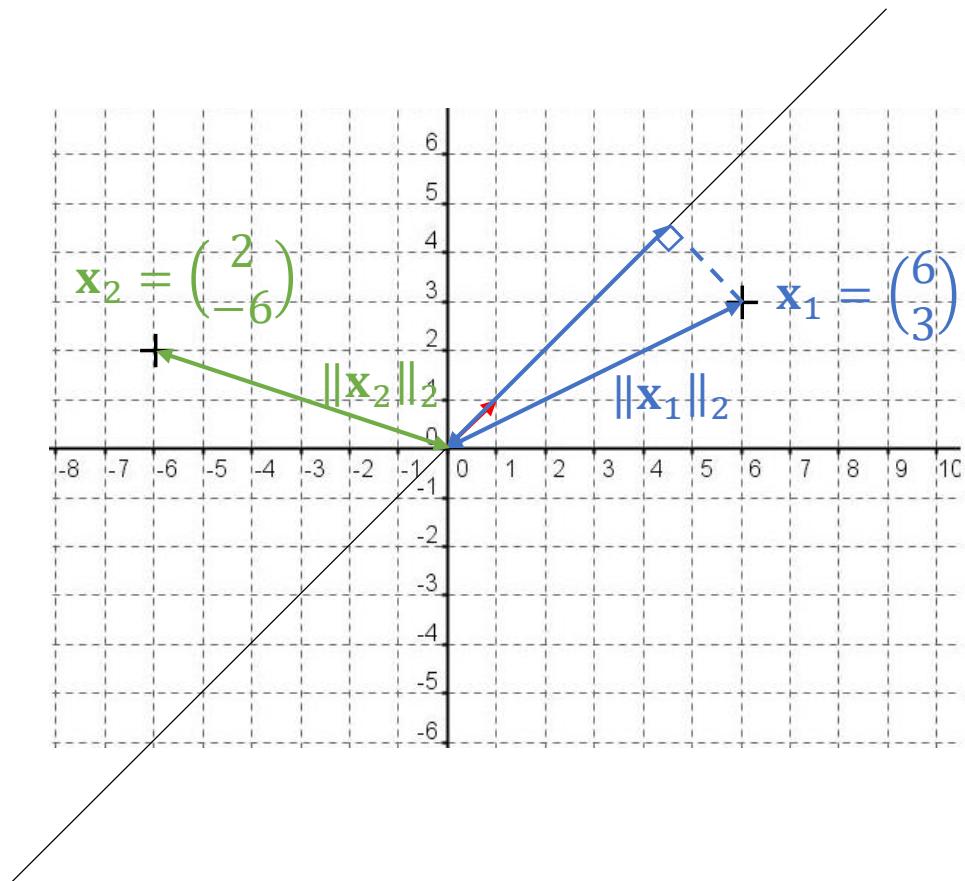


$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.



Total Variance – Example in 2D

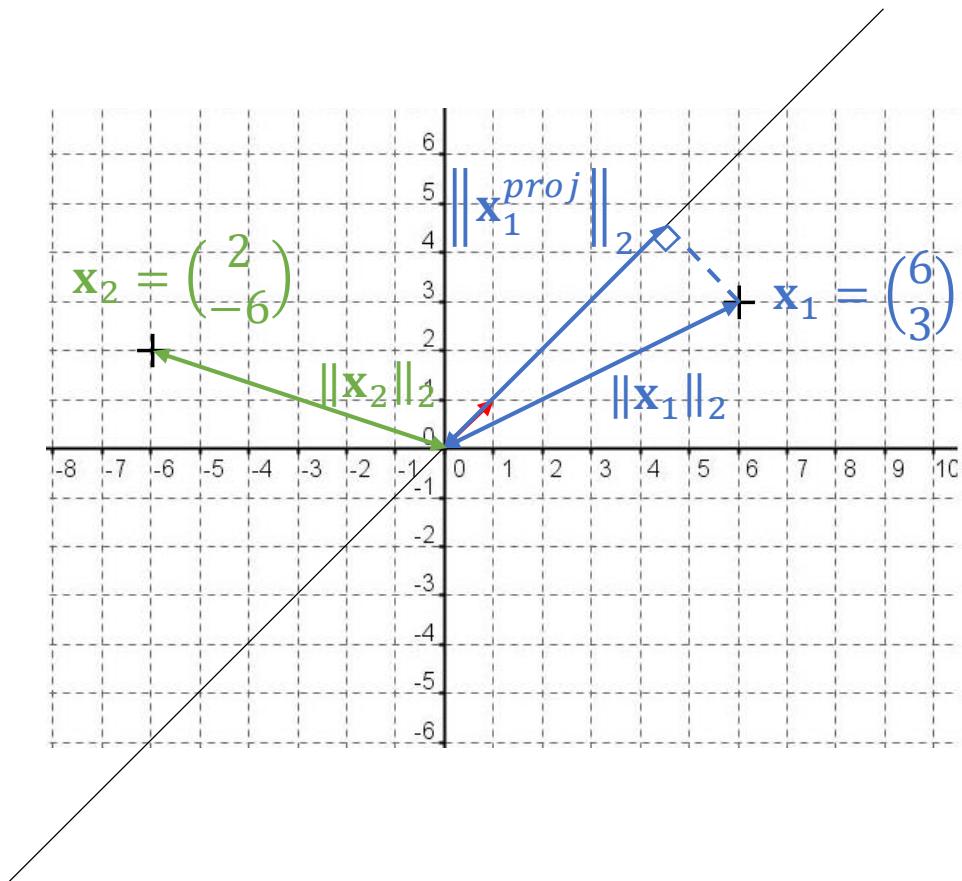


$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.



Total Variance – Example in 2D

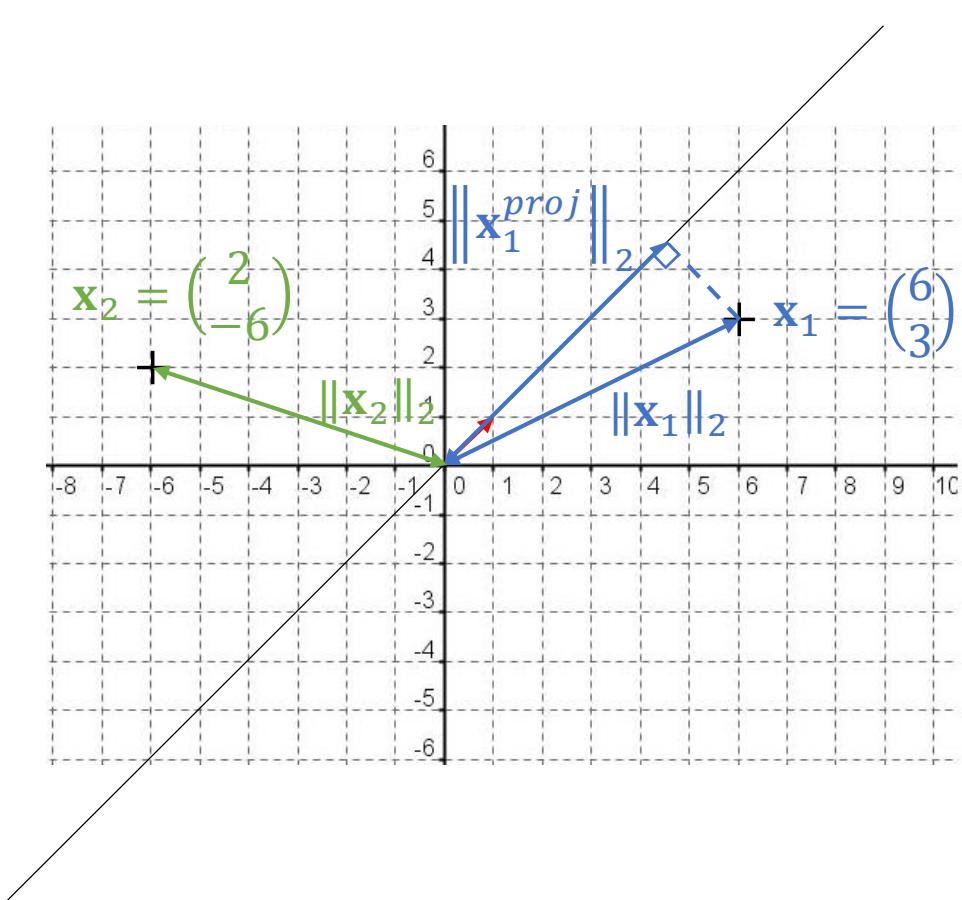


$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.



Total Variance – Example in 2D

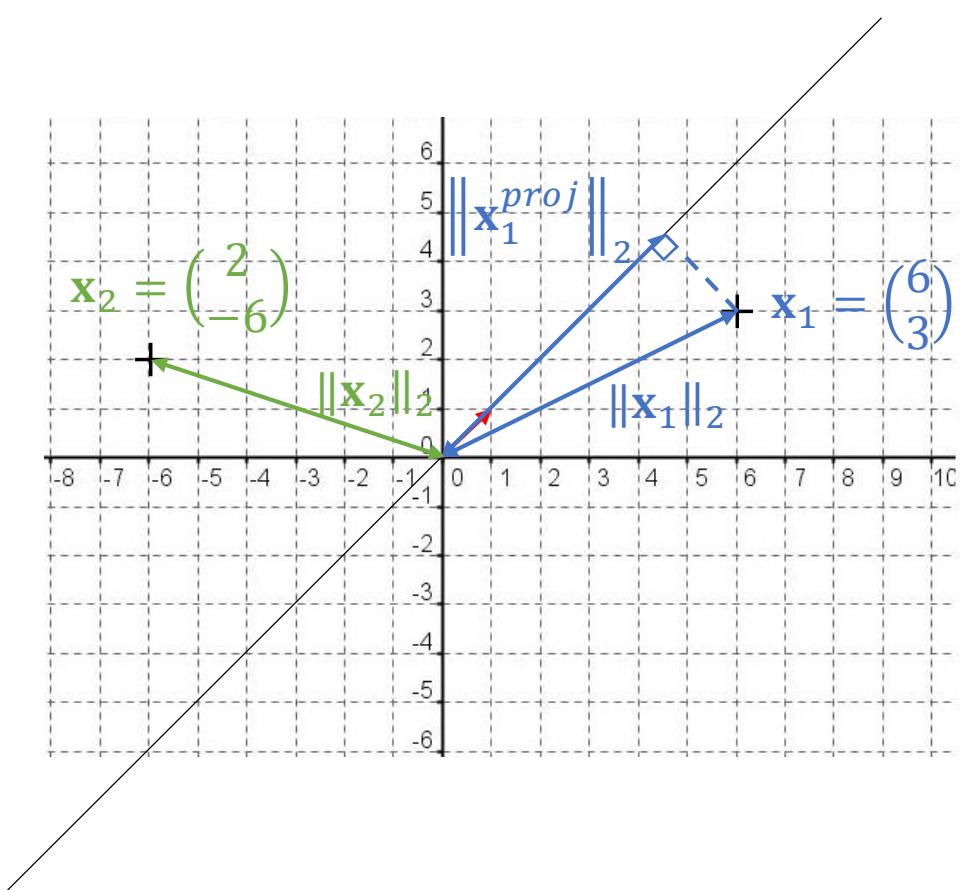


$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.



Total Variance – Example in 2D



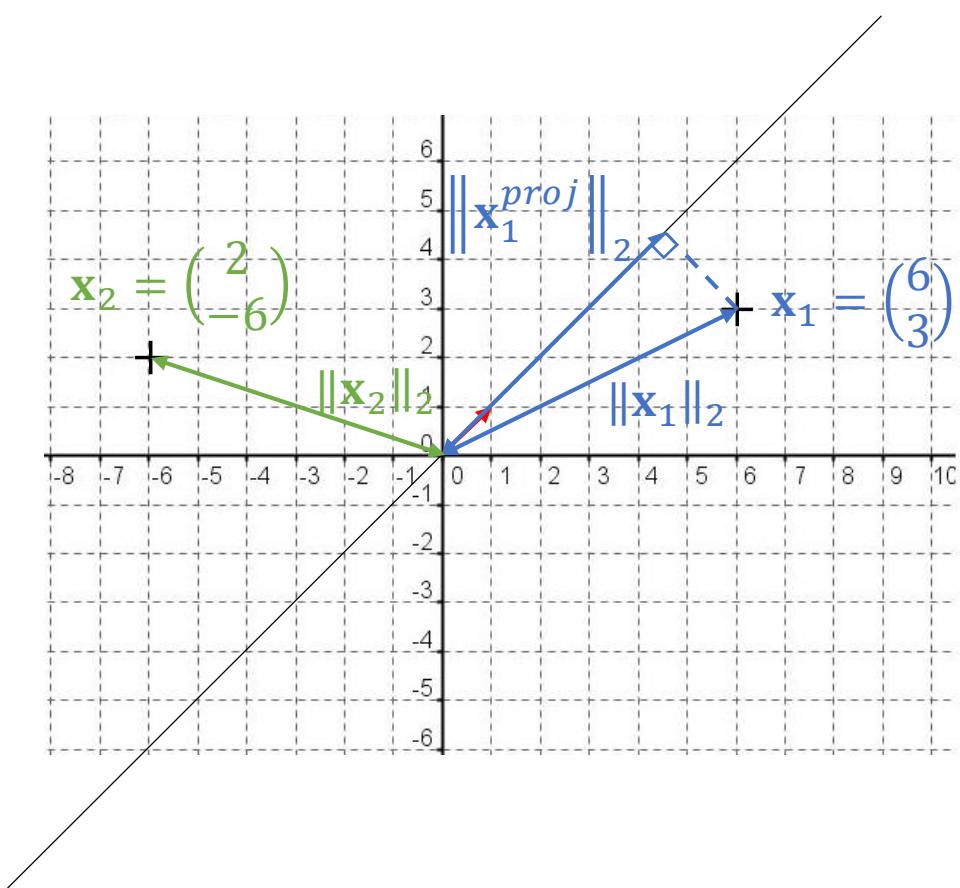
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \cdot \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} = \frac{1}{\sqrt{2}}(x_{11} + x_{12})$$



Total Variance – Example in 2D



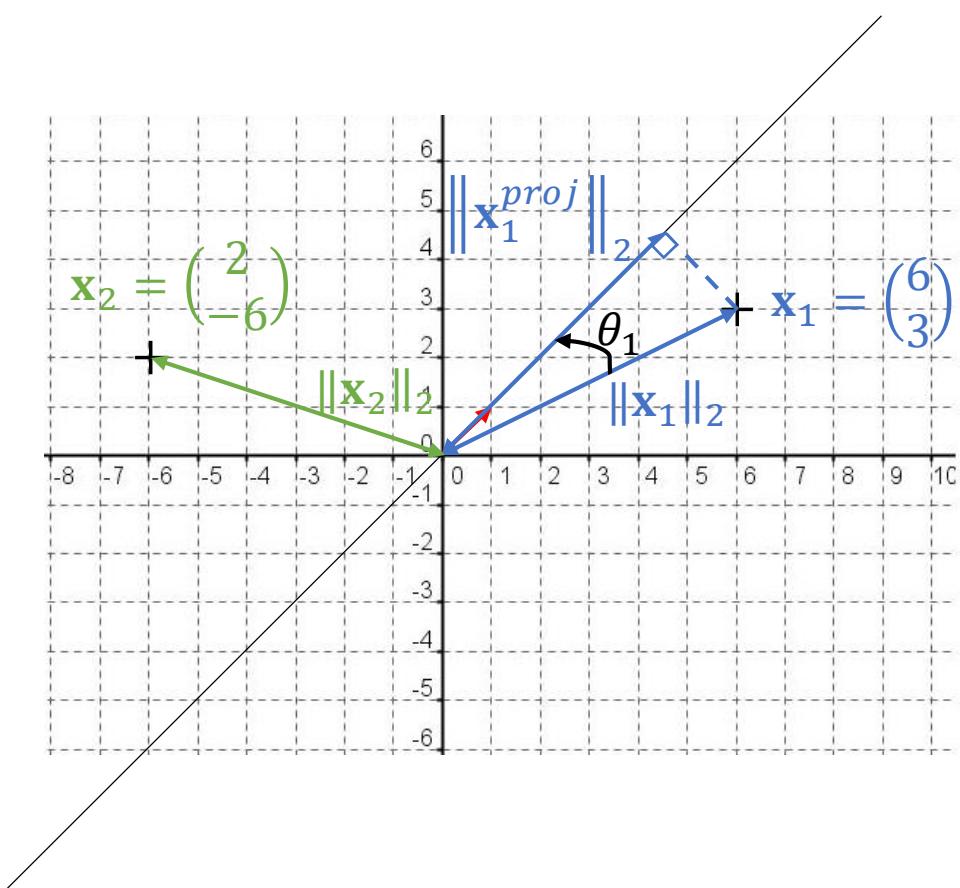
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$



Total Variance – Example in 2D



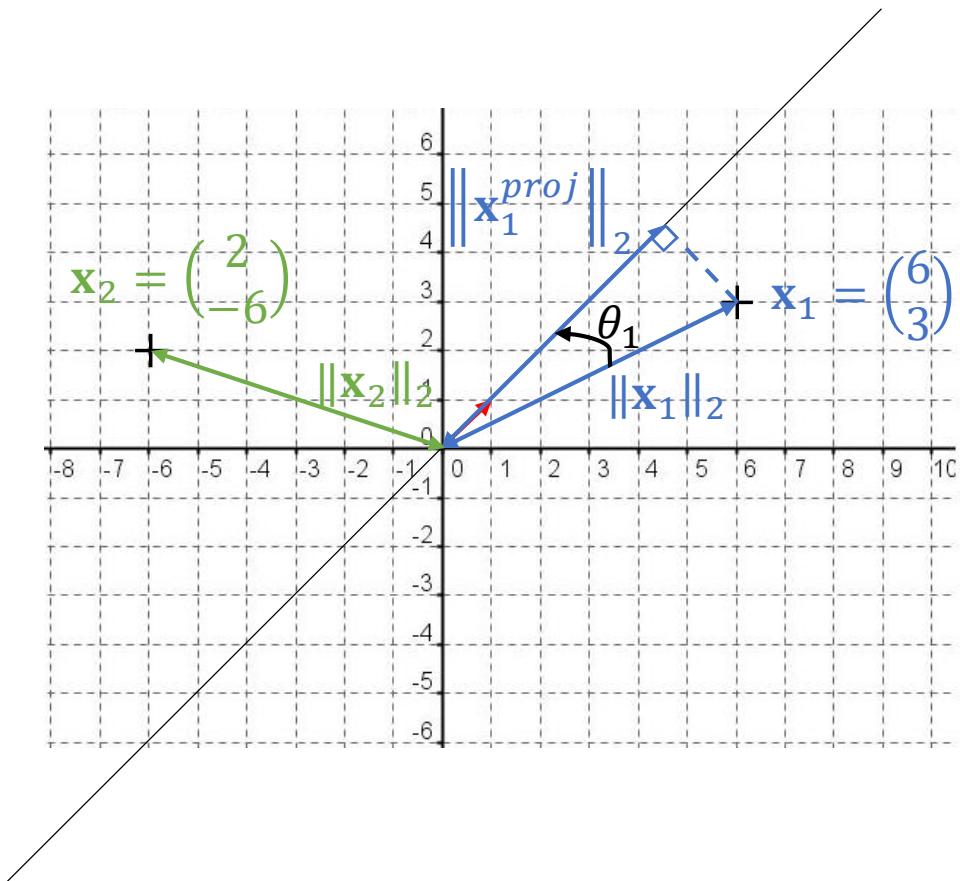
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\begin{aligned} \mathbf{x}_1^\top \mathbf{w} &= \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}} \\ &= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) \end{aligned}$$



Total Variance – Example in 2D



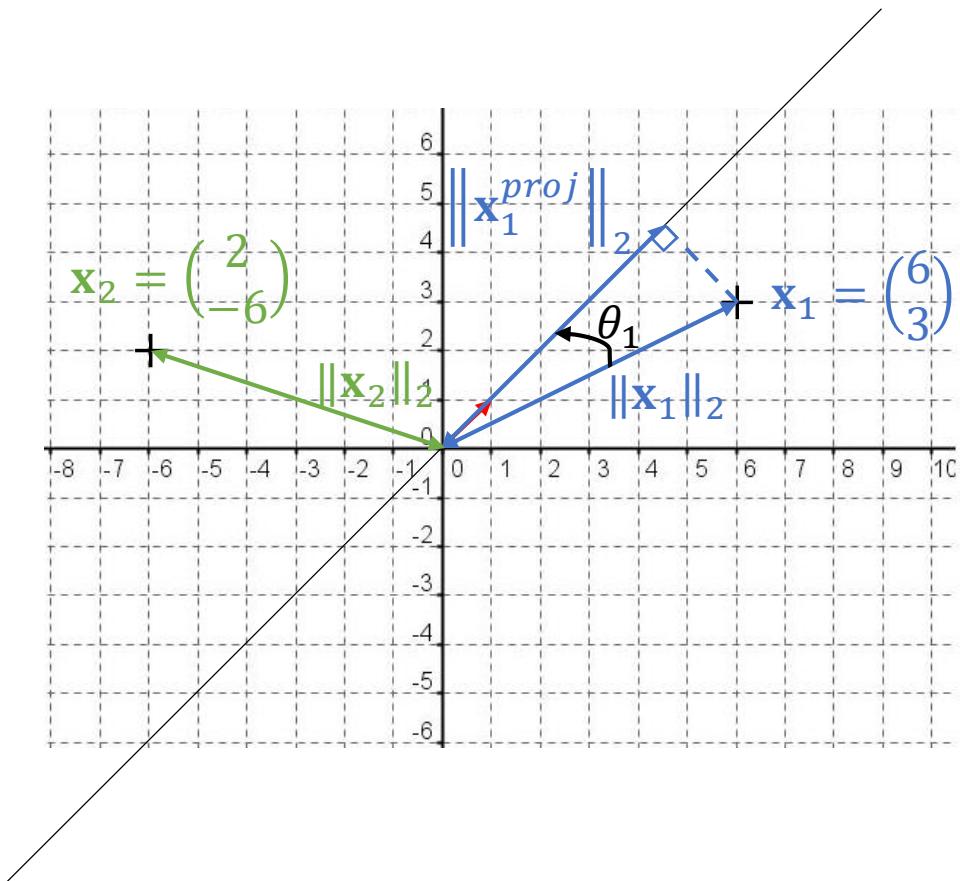
$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\begin{aligned} \mathbf{x}_1^\top \mathbf{w} &= \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}} \\ &= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) \end{aligned}$$



Total Variance – Example in 2D



$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

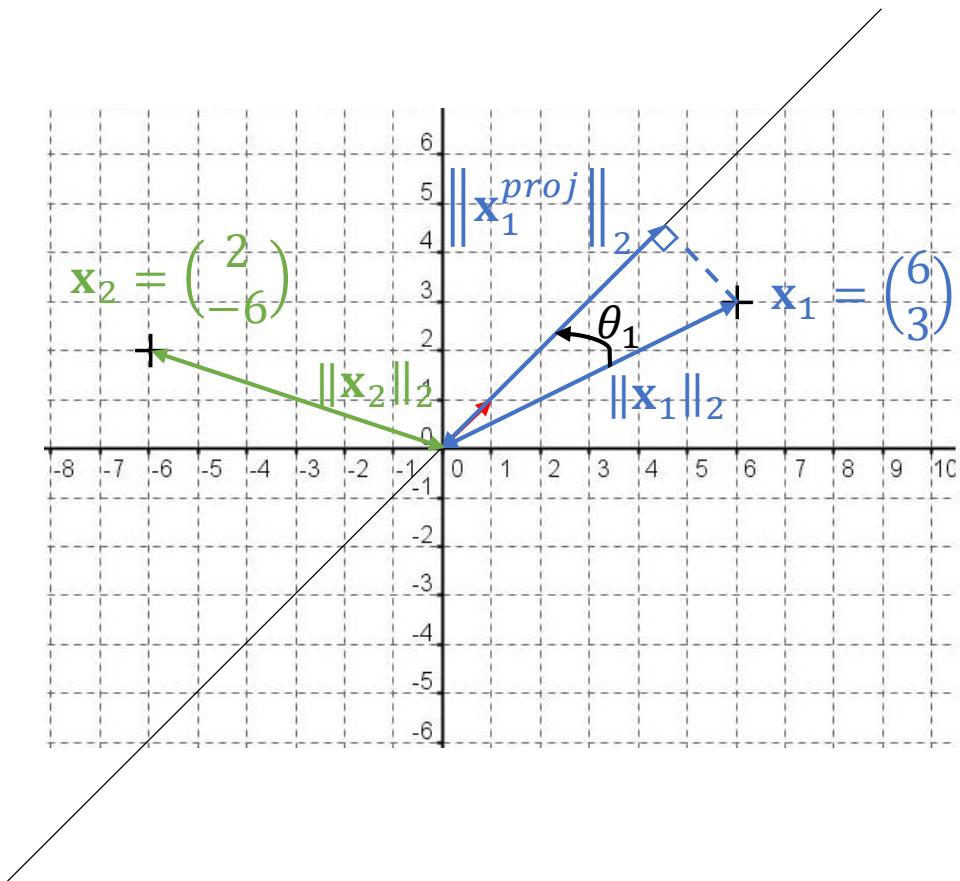
Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$

$$= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) = \|\mathbf{x}_1^{proj}\|_2$$



Total Variance – Example in 2D



$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

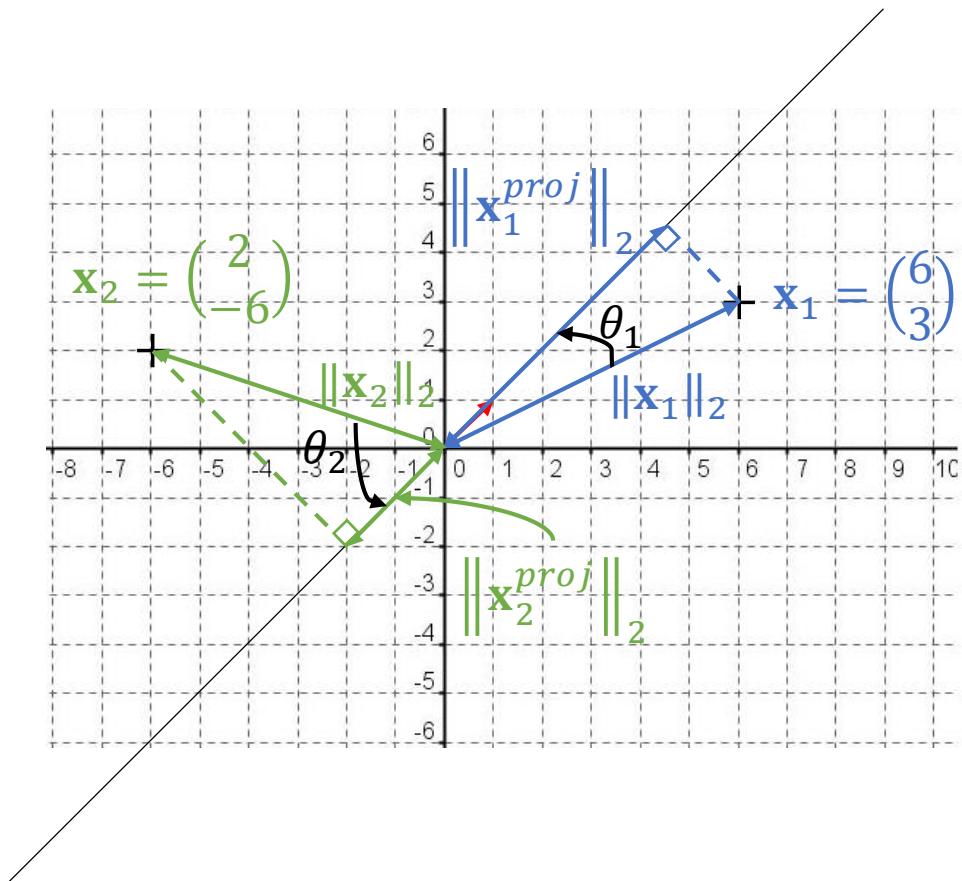
$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$

$$= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) = \|\mathbf{x}_1^{proj}\|_2$$

Similarly:



Total Variance – Example in 2D



$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$

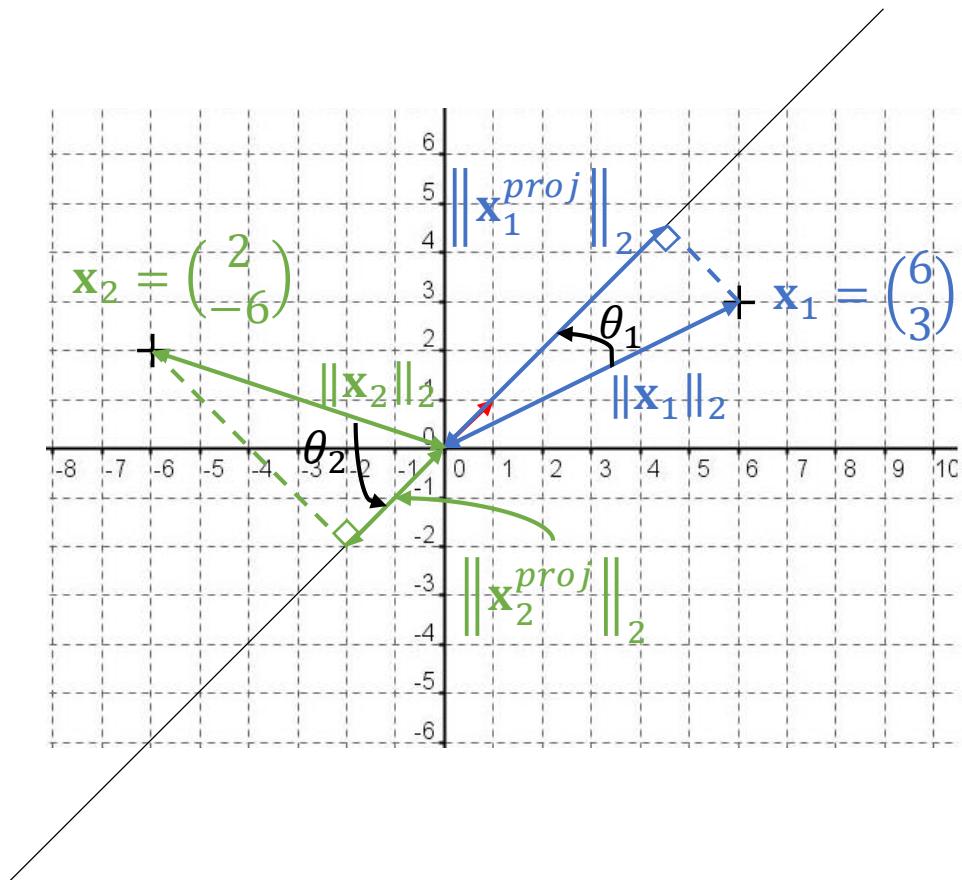
$$= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) = \|\mathbf{x}_1^{proj}\|_2$$

Similarly:

$$\mathbf{x}_2^\top \mathbf{w} = \frac{x_{21}}{\sqrt{2}} + \frac{x_{22}}{\sqrt{2}} = \|\mathbf{x}_2^{proj}\|_2$$



Total Variance – Example in 2D



$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$

$$= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) = \|\mathbf{x}_1^{\text{proj}}\|_2$$

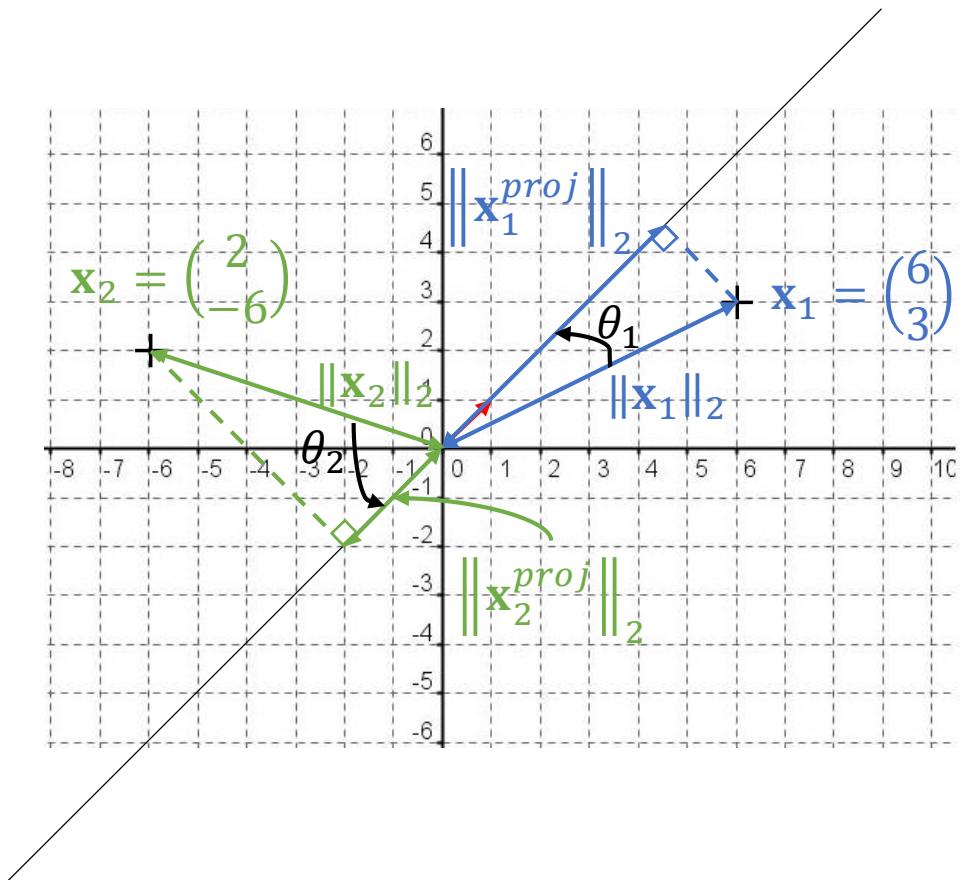
Similarly:

$$\mathbf{x}_2^\top \mathbf{w} = \frac{x_{21}}{\sqrt{2}} + \frac{x_{22}}{\sqrt{2}} = \|\mathbf{x}_2^{\text{proj}}\|_2$$

In the end:



Total Variance – Example in 2D



$$\|\mathbf{x}_1\|_2 = \sqrt{x_{11}^2 + x_{12}^2} \quad \|\mathbf{x}_2\|_2 = \sqrt{x_{21}^2 + x_{22}^2}$$

Director vector $\mathbf{w} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. We can see that $\|\mathbf{w}\|_2 = 1$.

$$\mathbf{x}_1^\top \mathbf{w} = \begin{pmatrix} x_{11} & x_{12} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \frac{x_{11}}{\sqrt{2}} + \frac{x_{12}}{\sqrt{2}}$$

$$= \|\mathbf{x}_1\|_2 \|\mathbf{w}\|_2 \cos(\theta_1) = \|\mathbf{x}_1\|_2 \cos(\theta_1) = \|\mathbf{x}_1^{\text{proj}}\|_2$$

Similarly:

$$\mathbf{x}_2^\top \mathbf{w} = \frac{x_{21}}{\sqrt{2}} + \frac{x_{22}}{\sqrt{2}} = \|\mathbf{x}_2^{\text{proj}}\|_2$$

In the end:

$$\text{TV}^{\text{proj}} = \frac{1}{2} ((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2)$$



PCA optimization criterion

$$\text{TV}^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right)$$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way:



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way: $\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way: $\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w}$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way:

$$\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way:

$$\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$$

Hence:



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way: $\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$

Hence:

$$TV^{\text{proj}} = \frac{1}{2} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way:

$$\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$$

Hence:

$$TV^{\text{proj}} = \frac{1}{2} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

It is possible to show in the general case that:



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way:

$$\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$$

Hence:

$$TV^{\text{proj}} = \frac{1}{2} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

It is possible to show in the general case that:

$$TV^{\text{proj}} = \frac{1}{n} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way: $\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$

Hence:

$$TV^{\text{proj}} = \frac{1}{2} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

It is possible to show in the general case that:

$$TV^{\text{proj}} = \frac{1}{n} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

Hence, one possible optimization criterion in order to estimate the first principal direction is:



PCA optimization criterion

$$TV^{\text{proj}} = \frac{1}{2} \left((\mathbf{x}_1^\top \mathbf{w})^2 + (\mathbf{x}_2^\top \mathbf{w})^2 \right) = \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} & \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix}$$

By the way: $\begin{pmatrix} \mathbf{x}_1^\top \mathbf{w} \\ \mathbf{x}_2^\top \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \end{pmatrix} \mathbf{w} = \mathbf{X}\mathbf{w}$

Hence:

$$TV^{\text{proj}} = \frac{1}{2} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

It is possible to show in the general case that:

$$TV^{\text{proj}} = \frac{1}{n} \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \text{Var}(\mathbf{X}\mathbf{w})$$

Hence, one possible optimization criterion in order to estimate the first principal direction is:

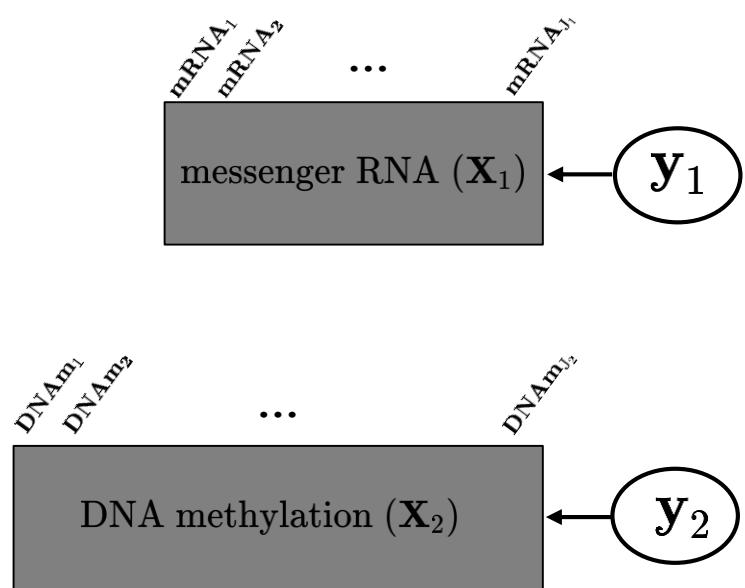
$$\max_{\mathbf{w}} \text{Var}(\mathbf{X}\mathbf{w})$$
$$\|\mathbf{w}\|_2^2 = 1$$



3. Unsupervised analysis with two-blocks

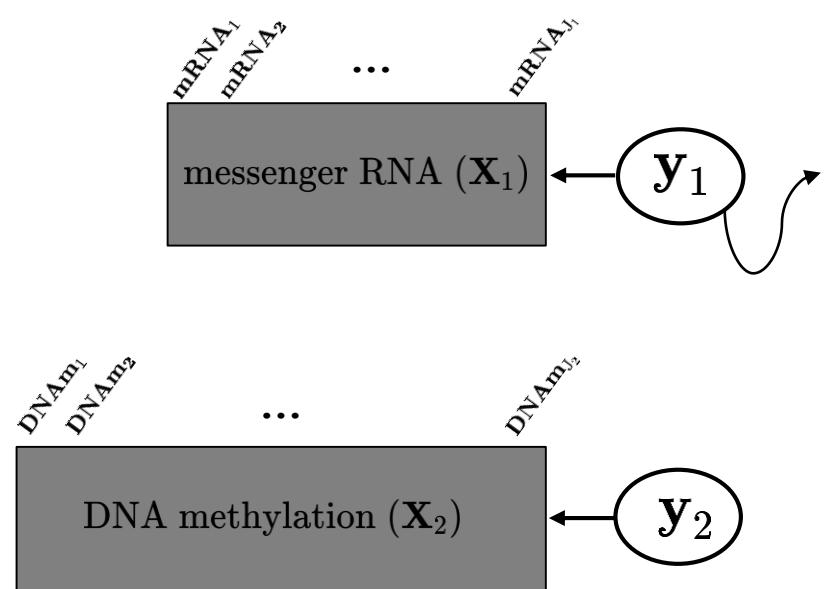


The philosophy of multiblock component methods



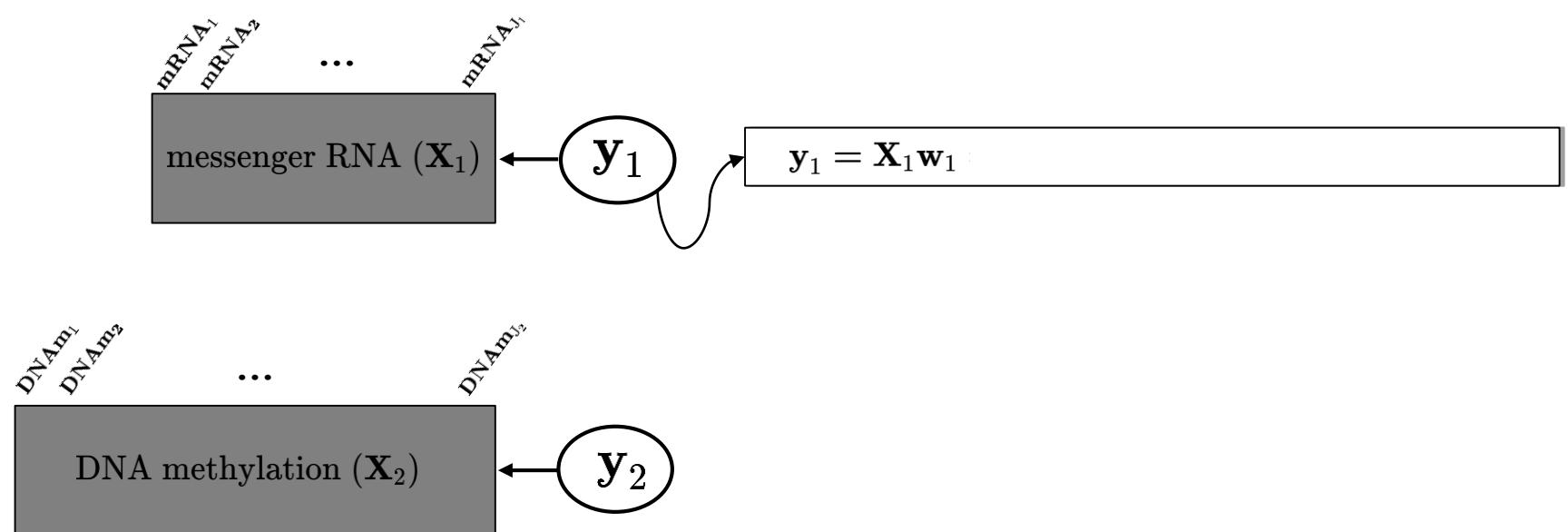


The philosophy of multiblock component methods



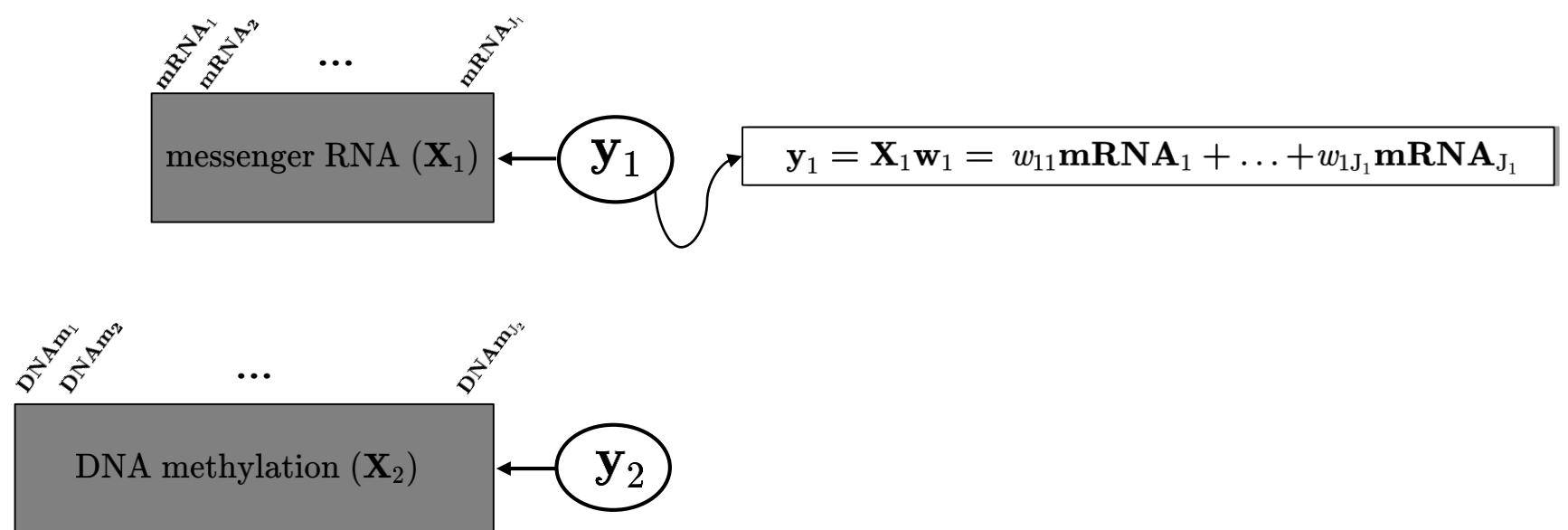


The philosophy of multiblock component methods



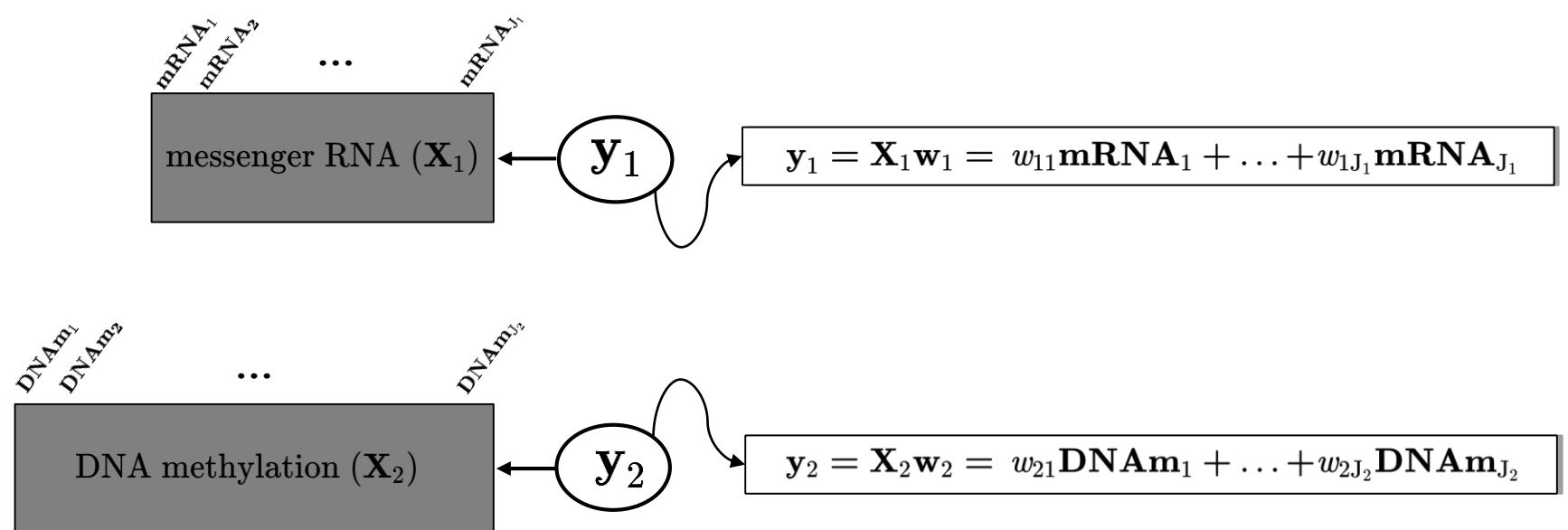


The philosophy of multiblock component methods



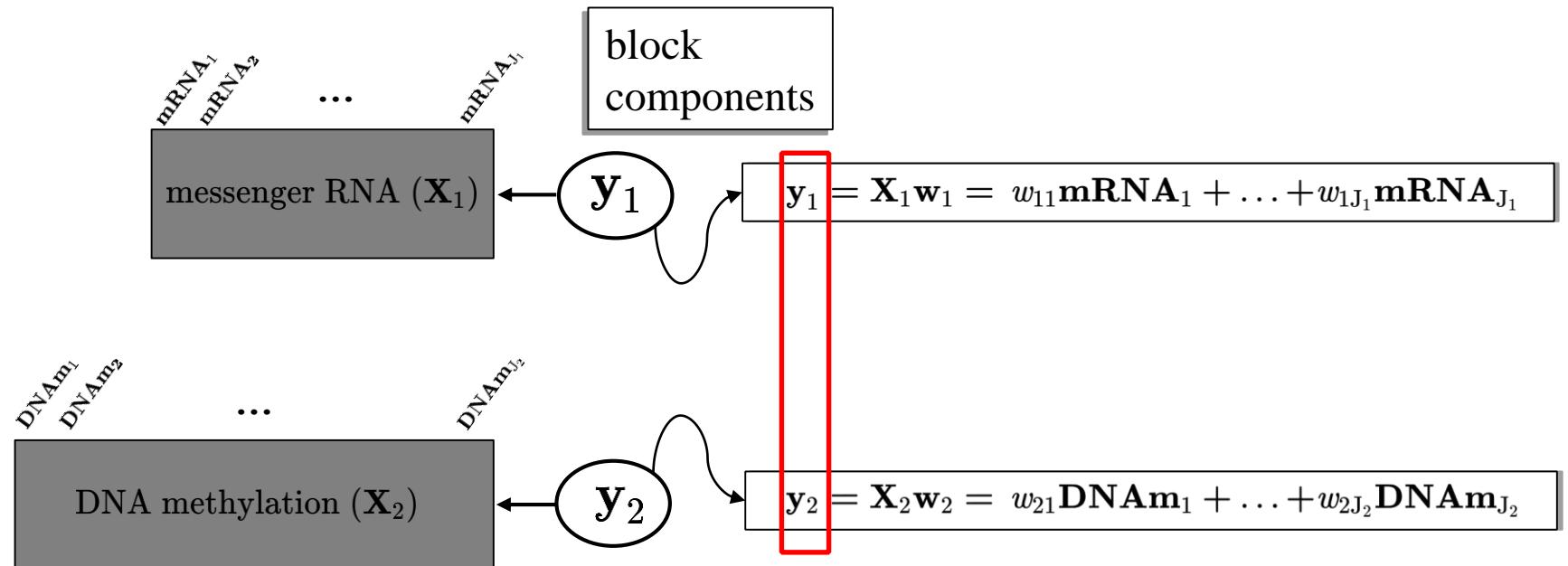


The philosophy of multiblock component methods



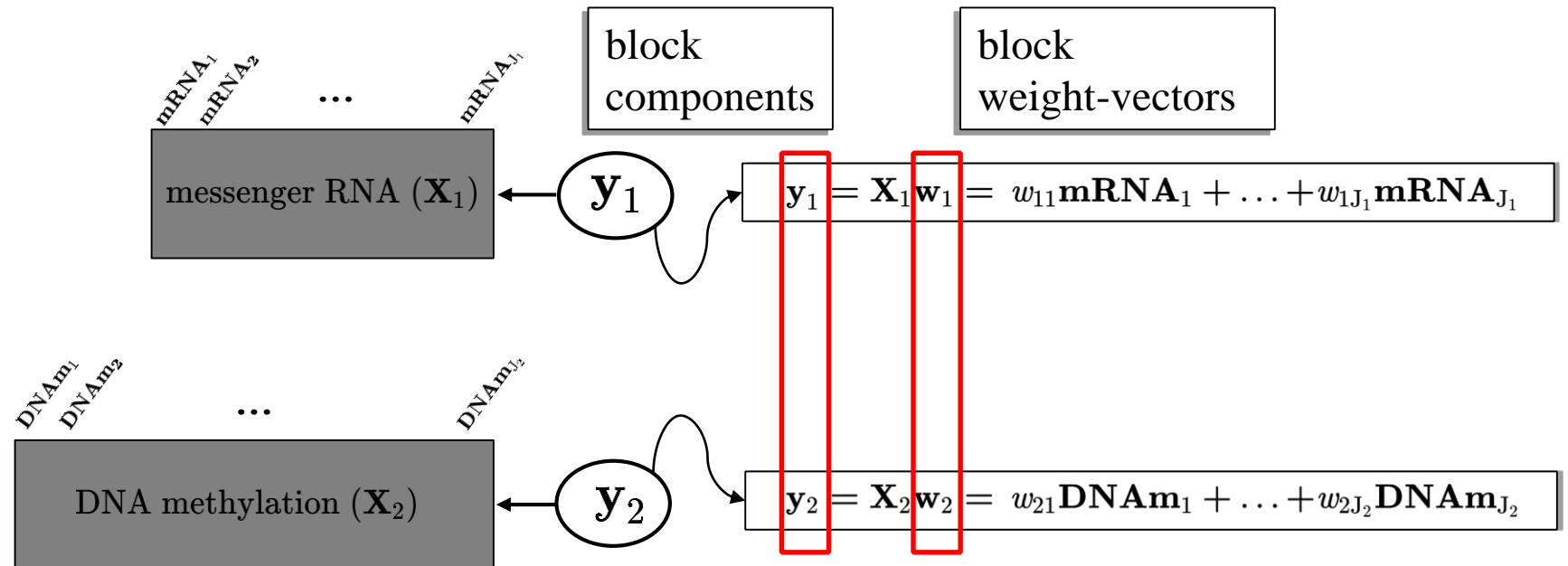


The philosophy of multiblock component methods

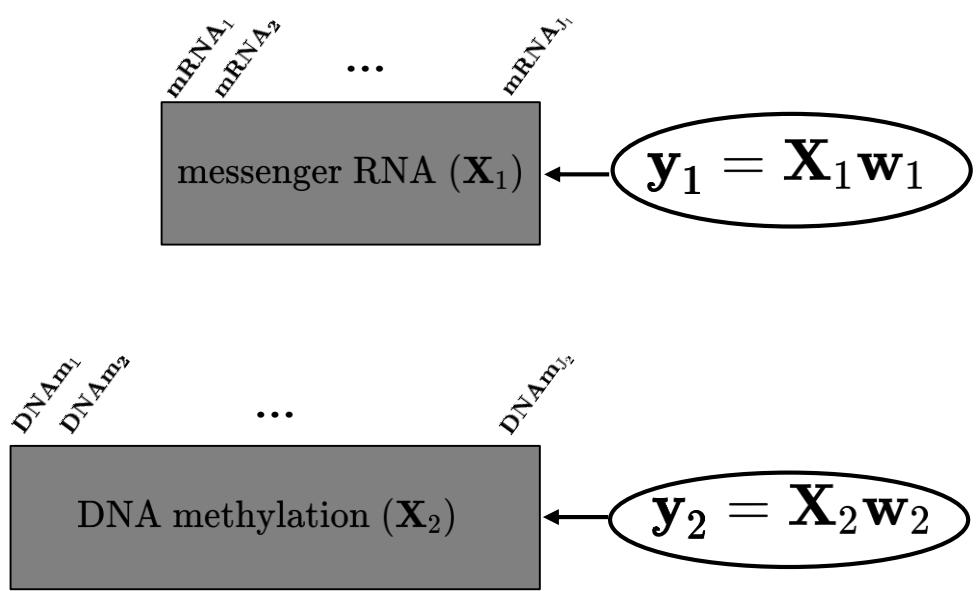




The philosophy of multiblock component methods



The philosophy of multiblock component methods

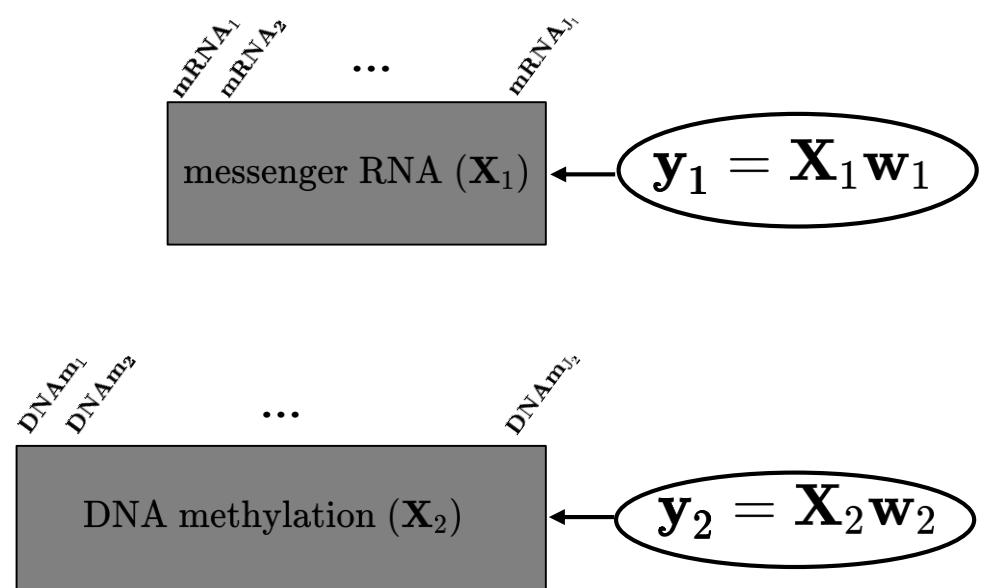


Block components should verify two properties at the same time:

1. Block components well explain their own block.
1. Block components are as correlated as possible for connected blocks.

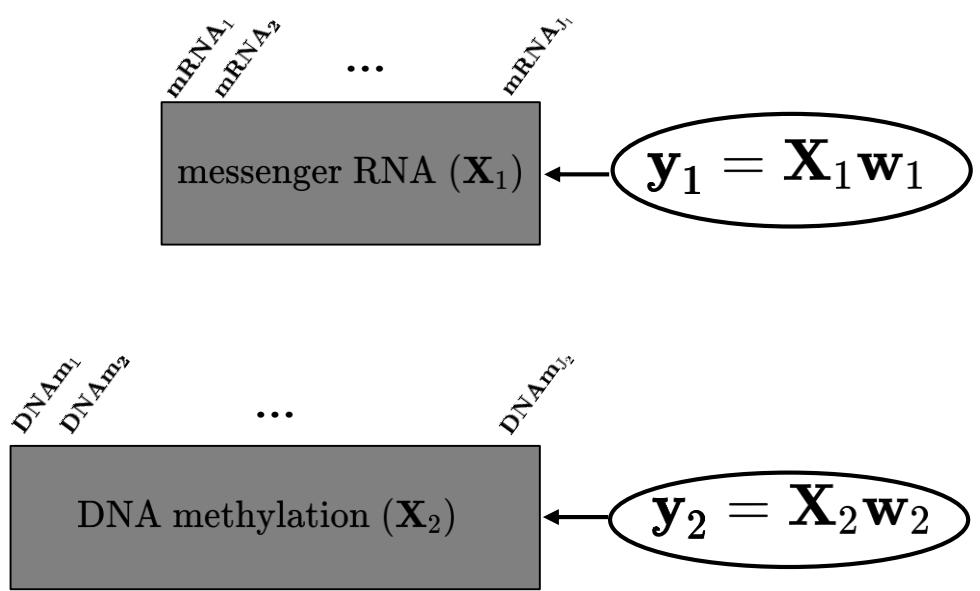


The philosophy of multiblock component methods





The philosophy of multiblock component methods



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

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

Courtesy to Arthur Tenenhaus.



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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

$$\text{Var}(\mathbf{X}_2 \mathbf{w}_2)$$



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA

Principal Component Analysis (PCA)

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



From PCA to PLS/CCA



From PCA to PLS/CCA

Partial Least Squares (PLS2)

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



From PCA to PLS/CCA

Canonical Correlation Analysis (CCA)

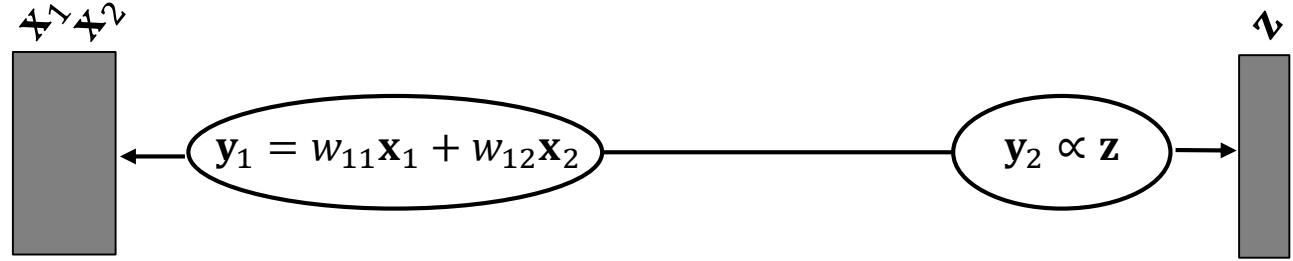
$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$
$$\text{Var}(\mathbf{X}_i \mathbf{w}_i) = 1$$

Partial Least Squares (PLS2)

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



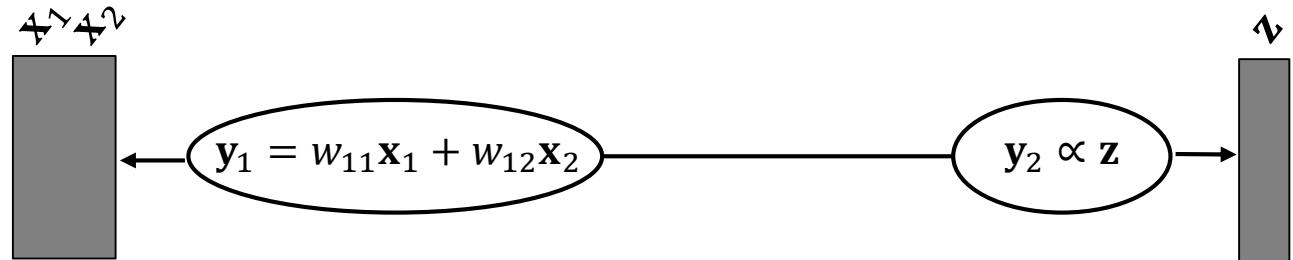
PLS & CCA with a figure





PLS & CCA with a figure

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

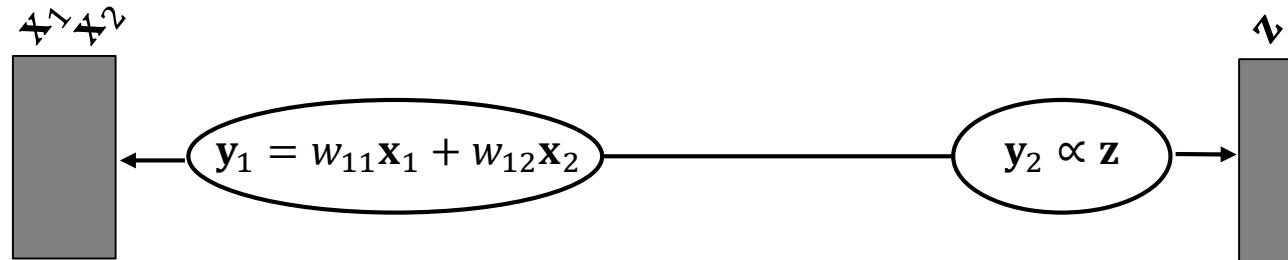




PLS & CCA with a figure

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

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

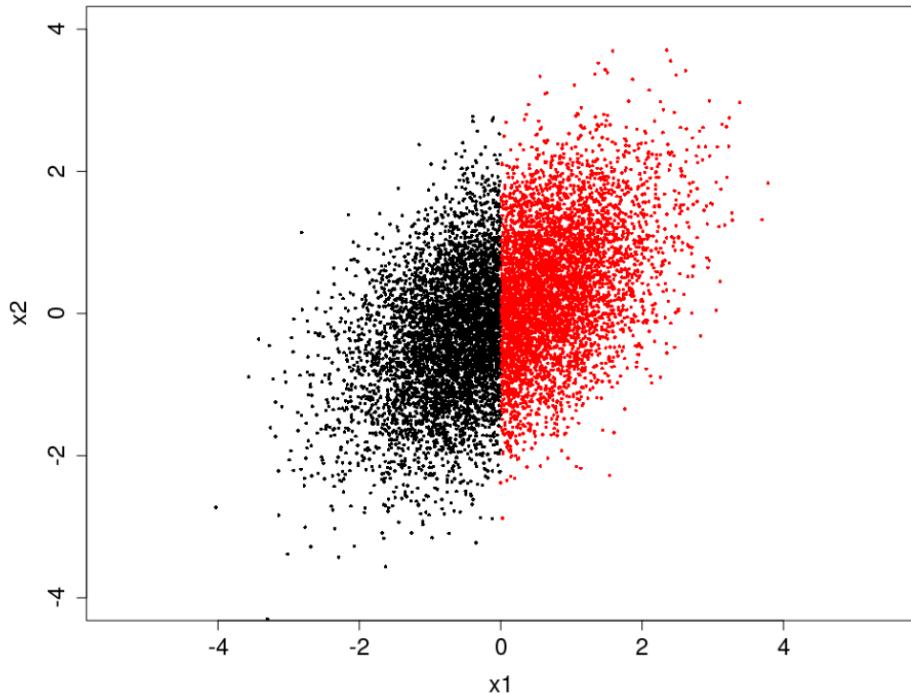
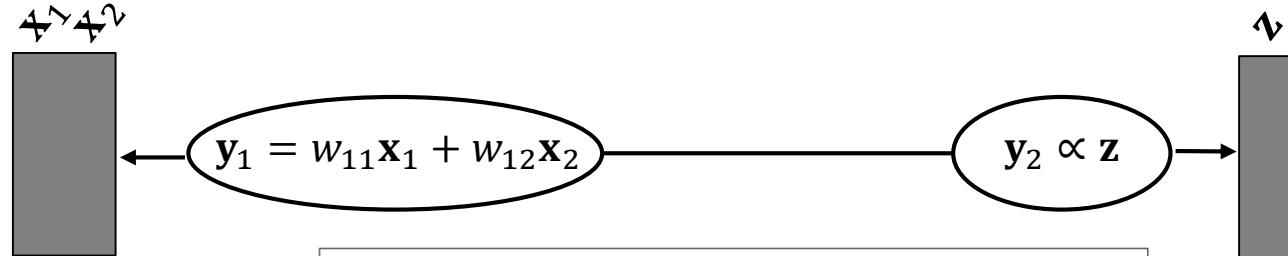




PLS & CCA with a figure

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

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

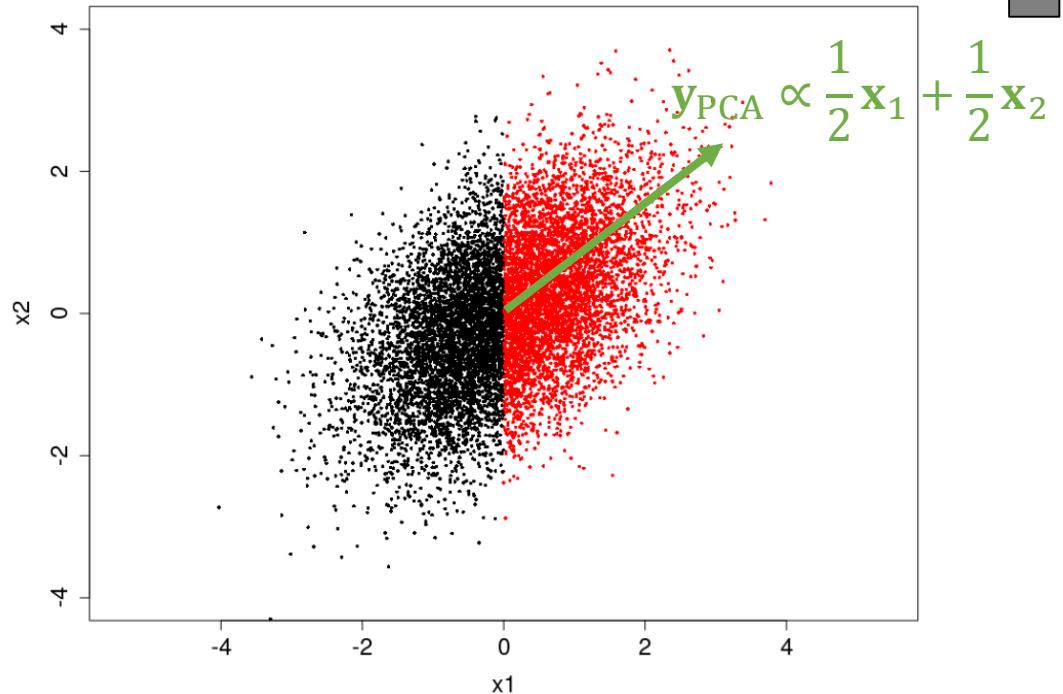
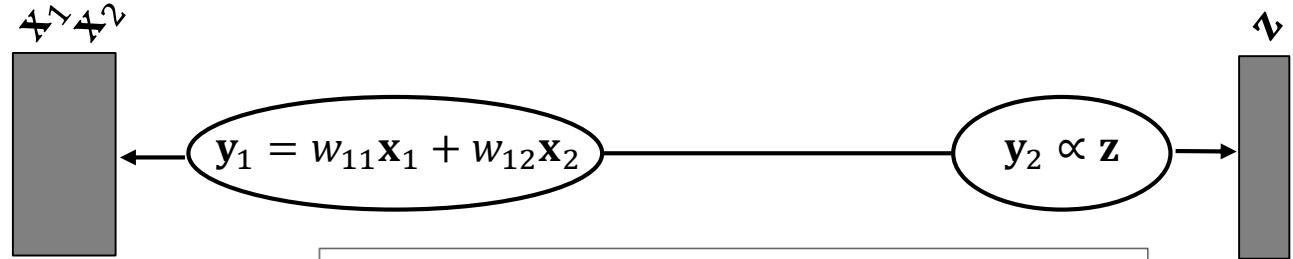




PLS & CCA with a figure

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

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

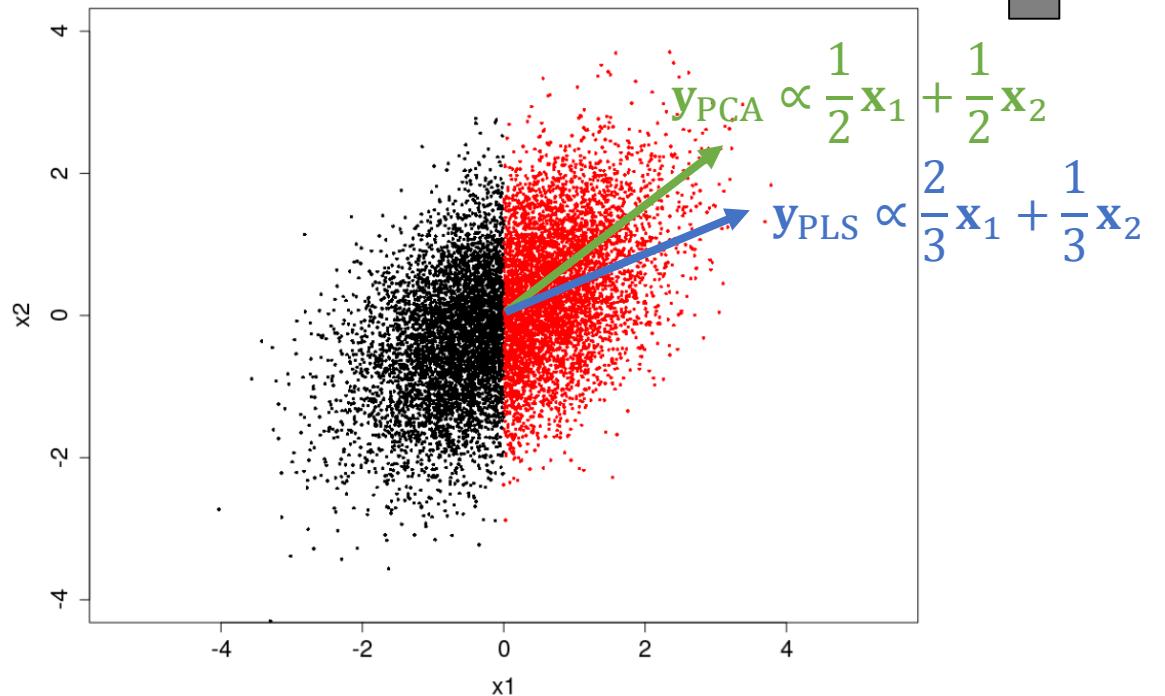
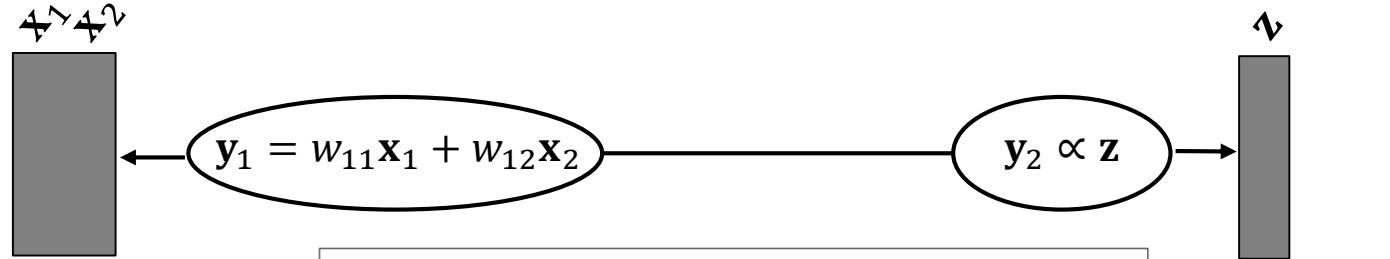




PLS & CCA with a figure

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

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

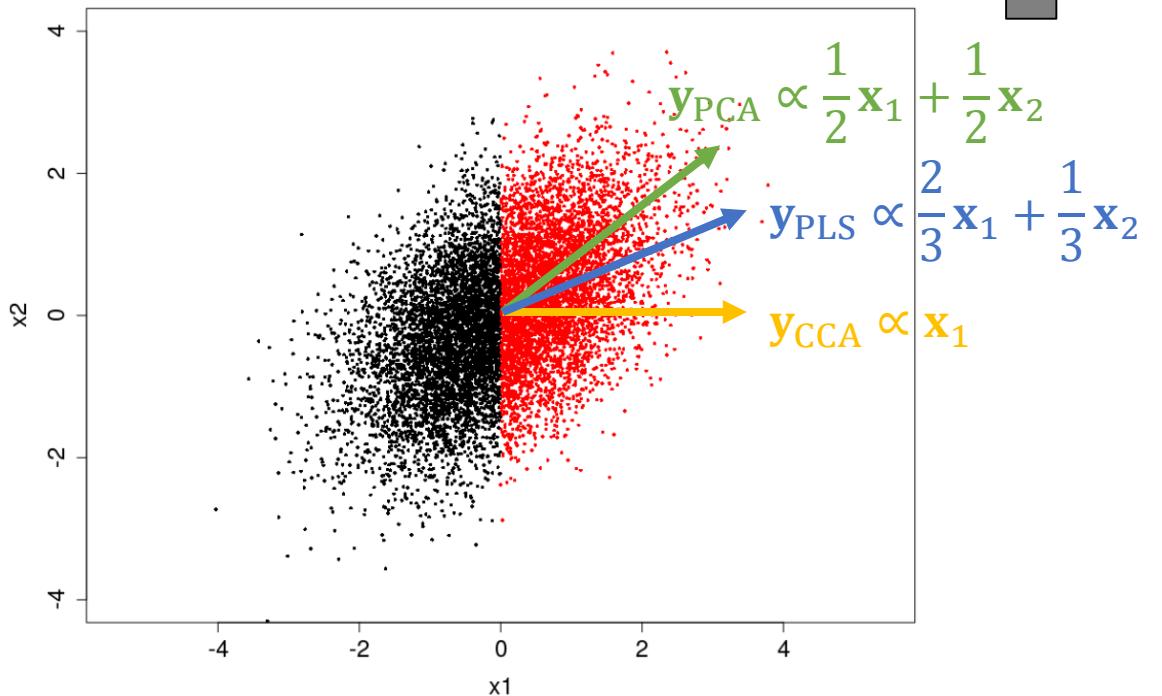
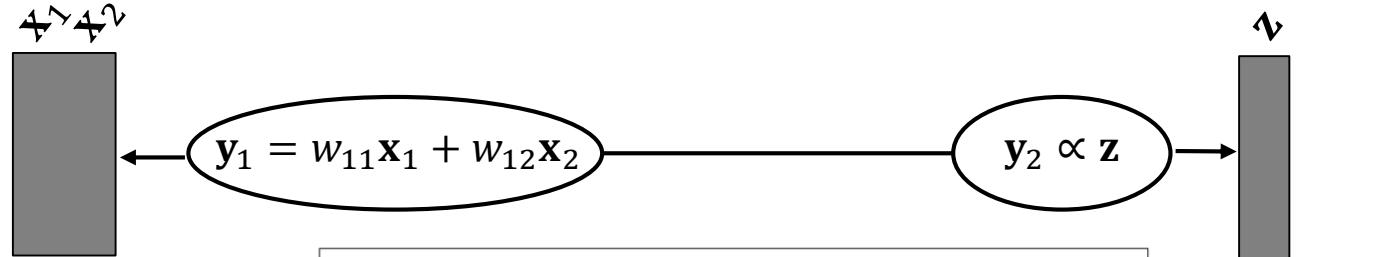




PLS & CCA with a figure

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

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





Two-blocks special cases: PLS & CCA

Canonical Correlation Analysis (CCA)

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

Partial Least Squares (PLS2)

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



Two-blocks special cases: PLS & CCA ... and Regularized-CCA

Canonical Correlation Analysis (CCA)

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

Partial Least Squares (PLS2)

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

Regularized-CCA

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

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



Two-blocks special cases: PLS & CCA ... and Regularized-CCA

Canonical Correlation Analysis (CCA)

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

$\text{Var}(\mathbf{X}_i \mathbf{w}_i) = 1$



Partial Least Squares (PLS2)

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

$\|\mathbf{w}_i\|_2^2 = 1$

Regularized-CCA

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

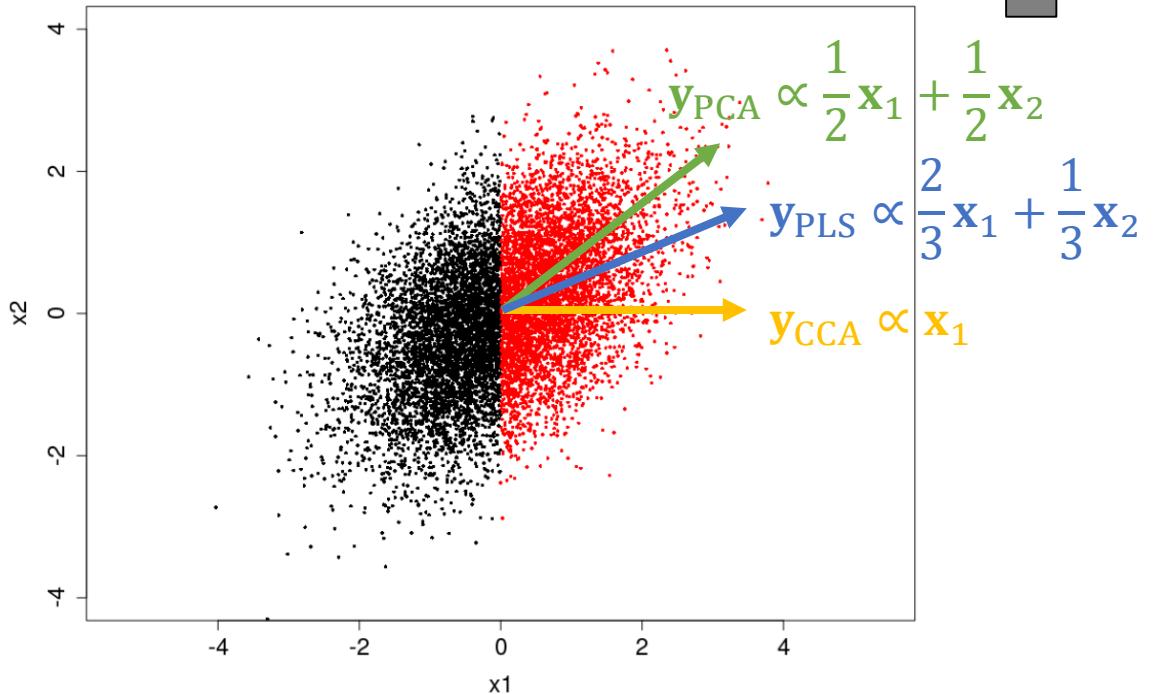
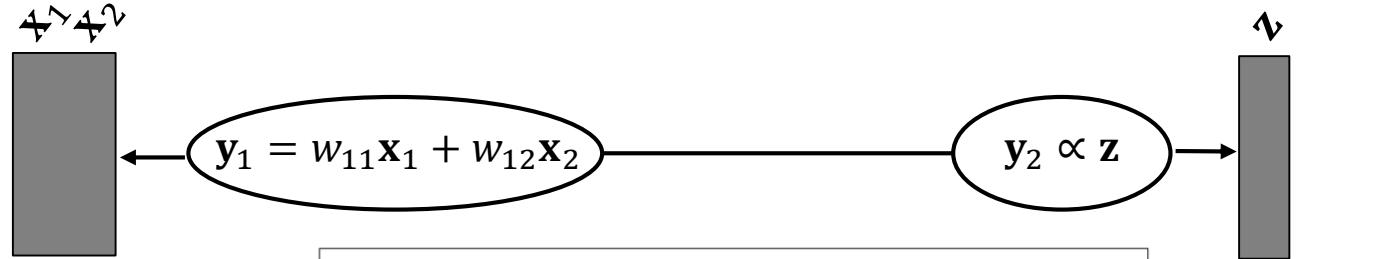
$$\text{s. t. } (1 - \tau_i) \text{Var}(\mathbf{X}_i \mathbf{w}_i) + \tau_i \|\mathbf{w}_i\|_2^2 = 1.$$



PLS & CCA with a figure

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

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

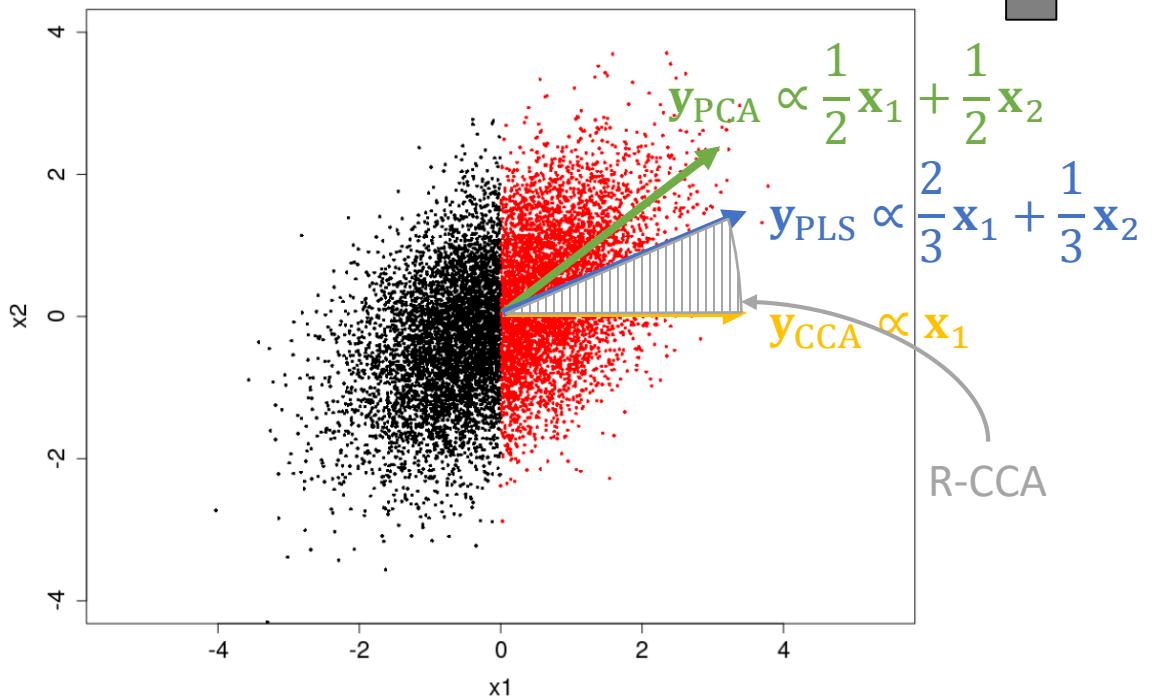
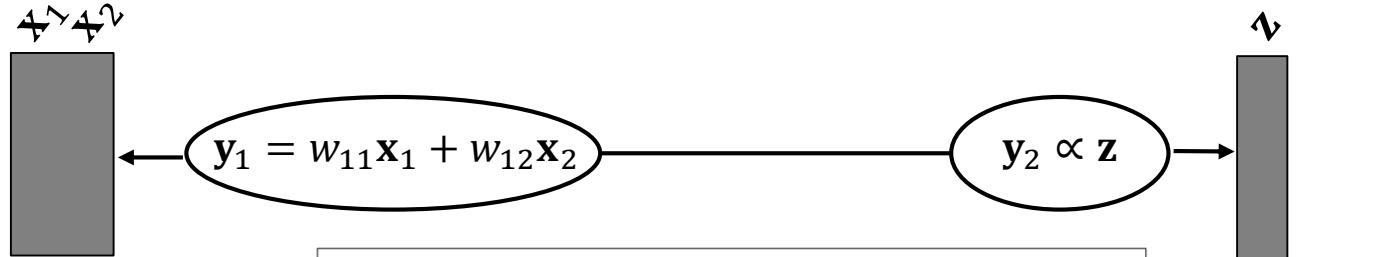




PLS & CCA with a figure

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

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

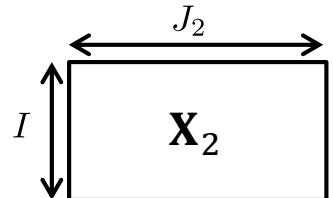
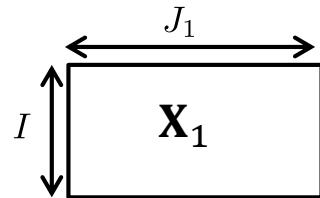




4. Unsupervised analysis with L -blocks

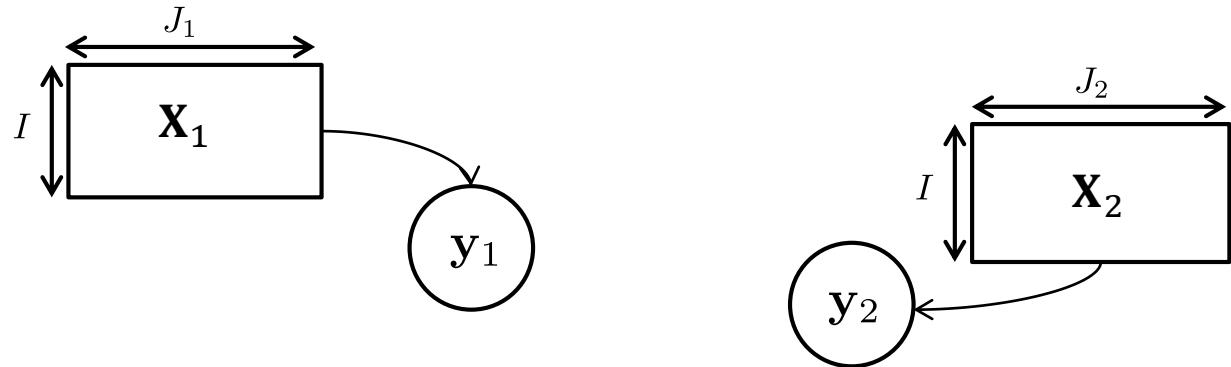


Regularized Generalized Canonical Correlation Analysis (RGCCA)



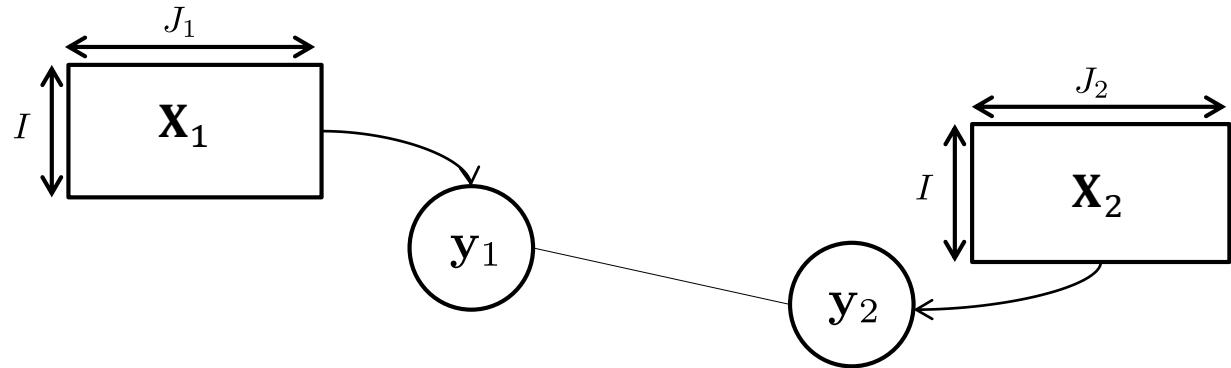


Regularized Generalized Canonical Correlation Analysis (RGCCA)



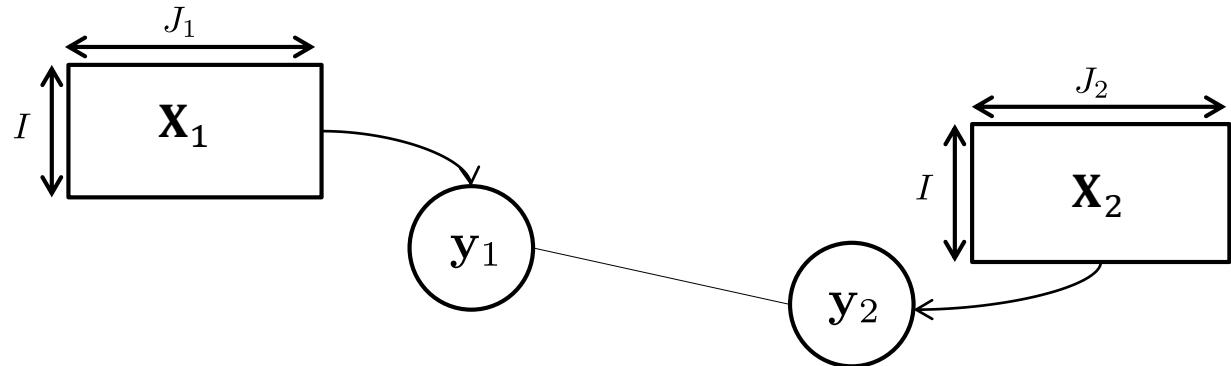


Regularized Generalized Canonical Correlation Analysis (RGCCA)





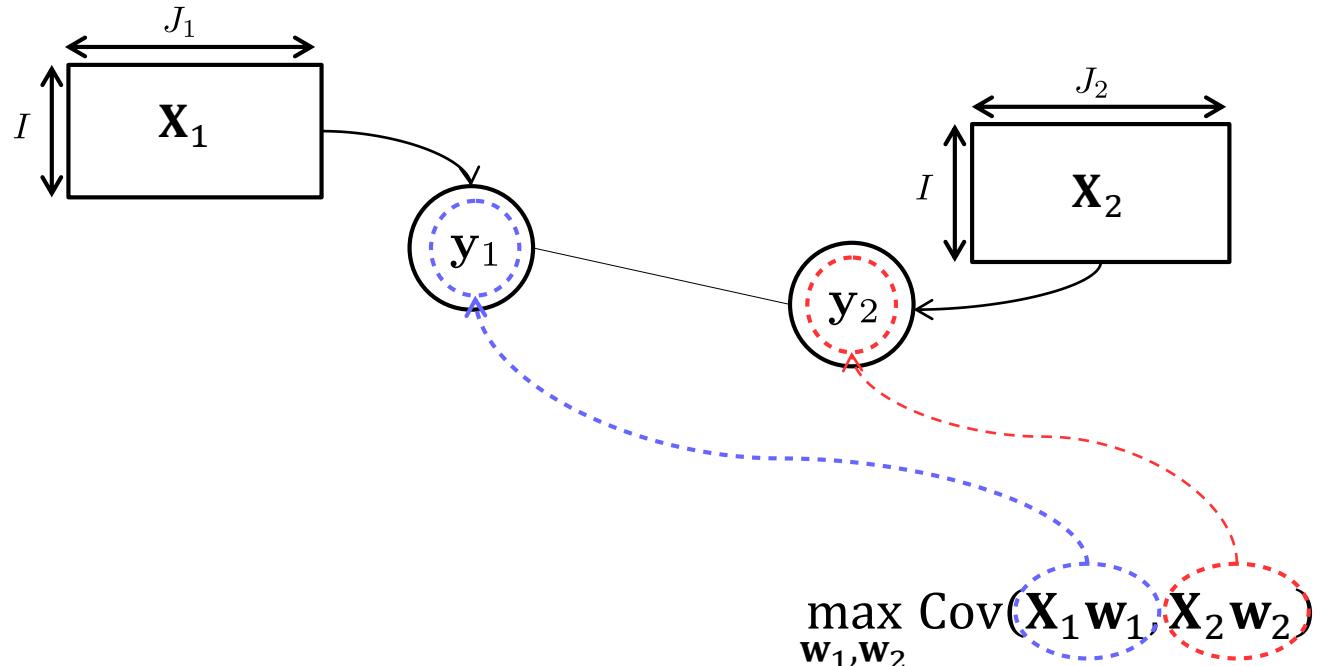
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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

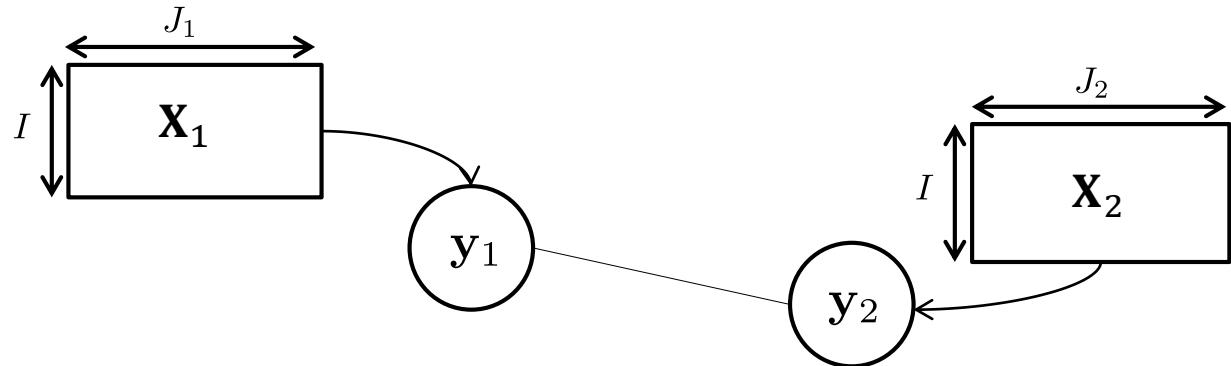


Regularized Generalized Canonical Correlation Analysis (RGCCA)





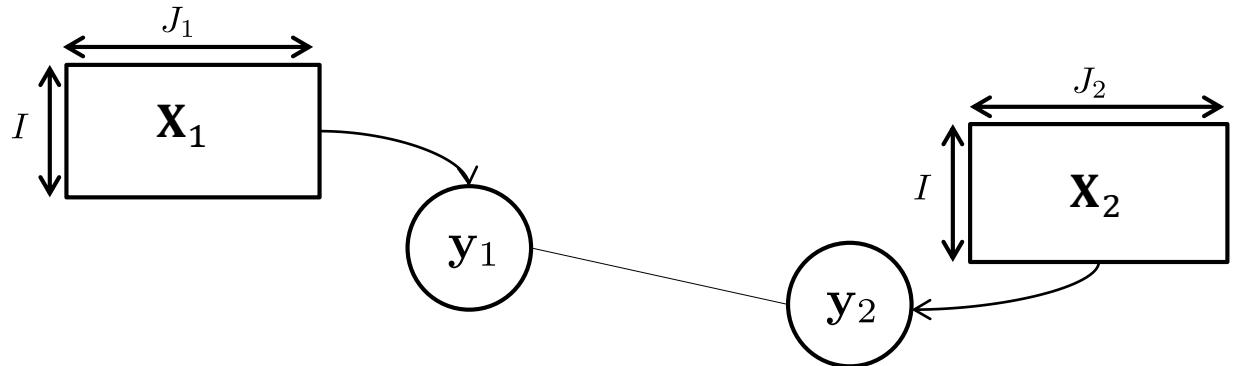
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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



Regularized Generalized Canonical Correlation Analysis (RGCCA)

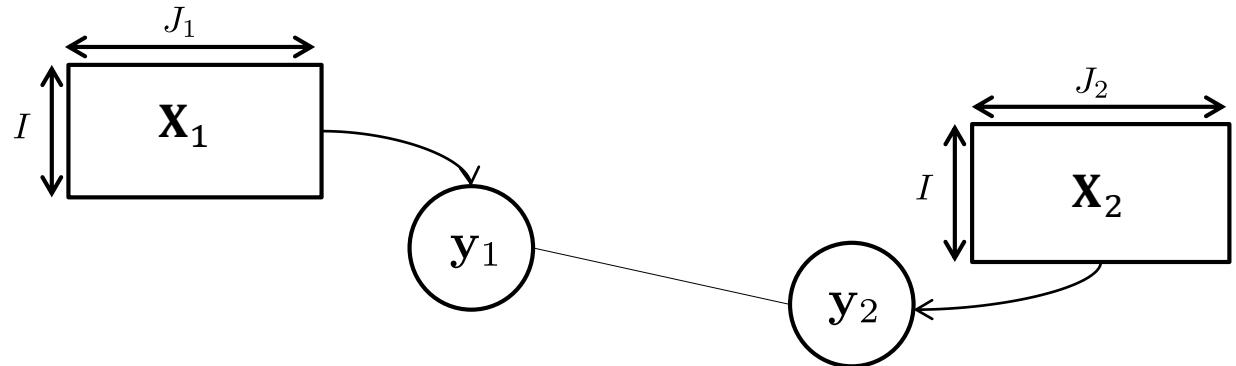


$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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



Regularized Generalized Canonical Correlation Analysis (RGCCA)



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

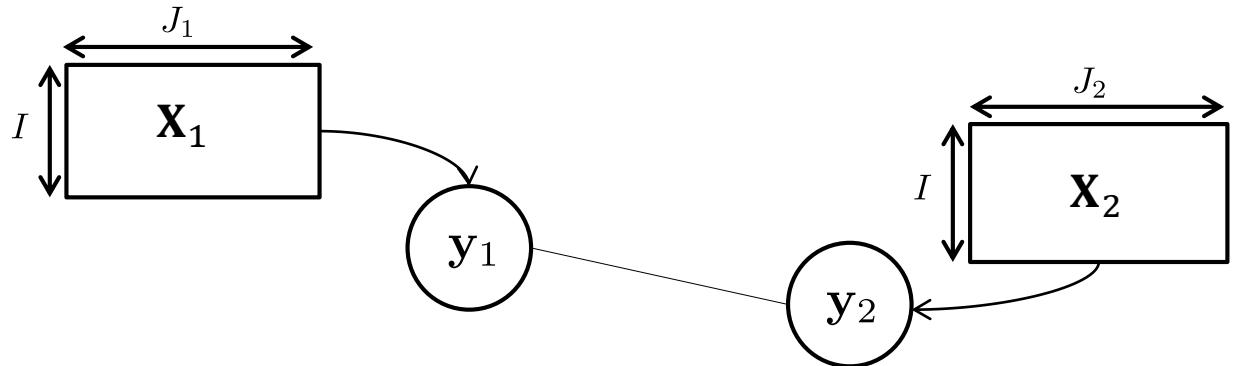
$$\text{s. t. } \mathbf{w}_l^\top \mathbf{X}_l^\top \mathbf{X}_l \mathbf{w}_l = I, \quad l = 1, 2.$$



Canonical Correlation Analysis



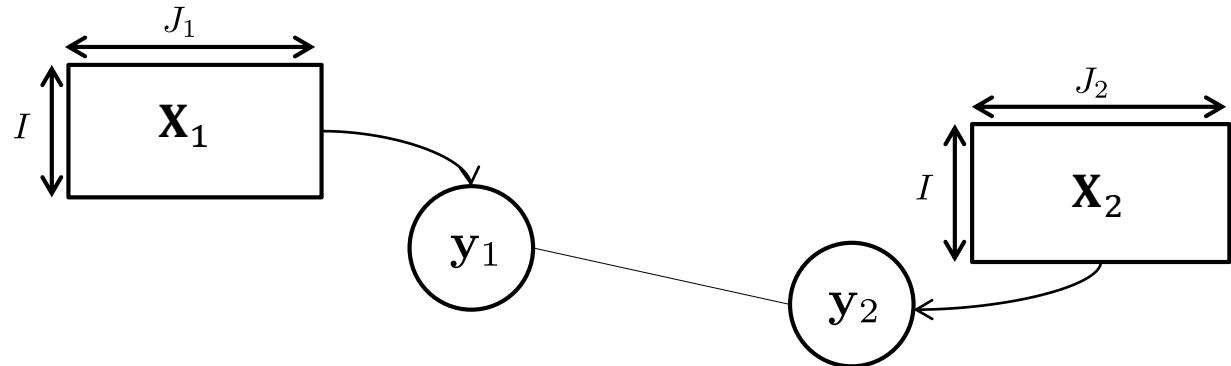
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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

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

Regularized Generalized Canonical Correlation Analysis (RGCCA)

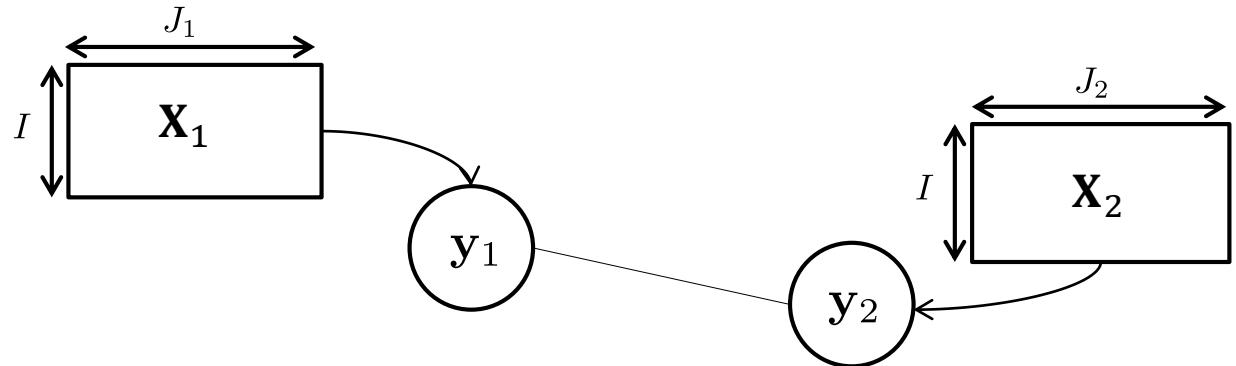


$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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



Regularized Generalized Canonical Correlation Analysis (RGCCA)



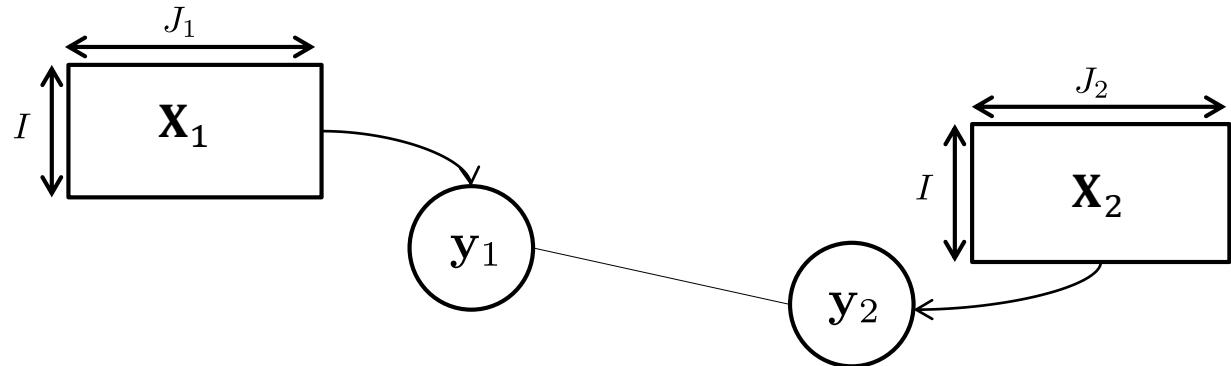
$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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



Partial Least Squares 2

Regularized Generalized Canonical Correlation Analysis (RGCCA)

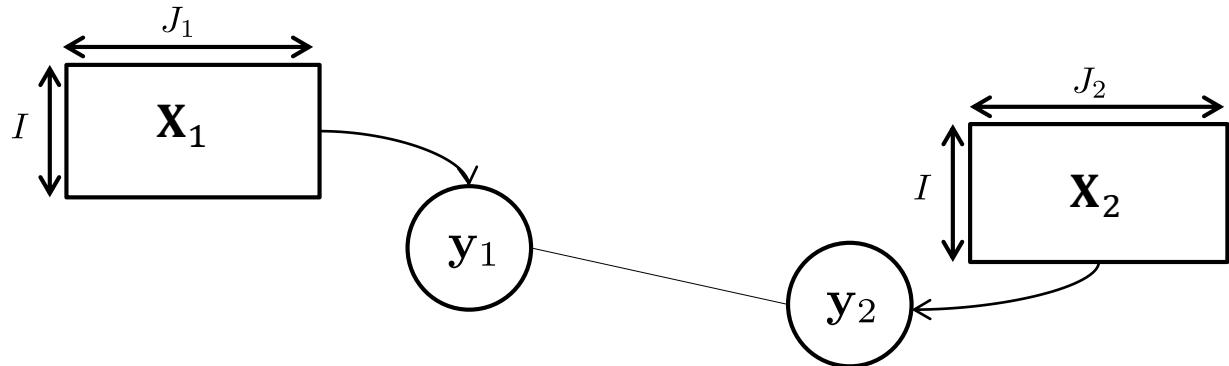


$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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



Regularized Generalized Canonical Correlation Analysis (RGCCA)

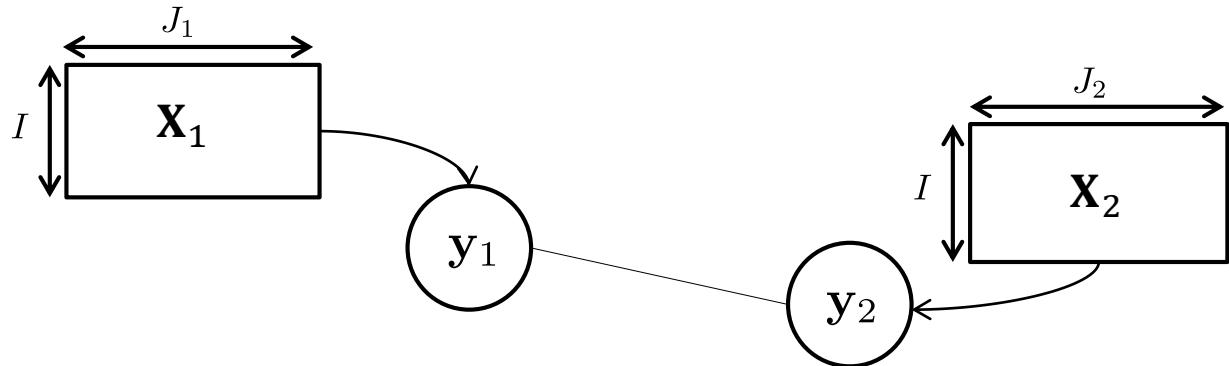


$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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



Regularized Generalized Canonical Correlation Analysis (RGCCA)



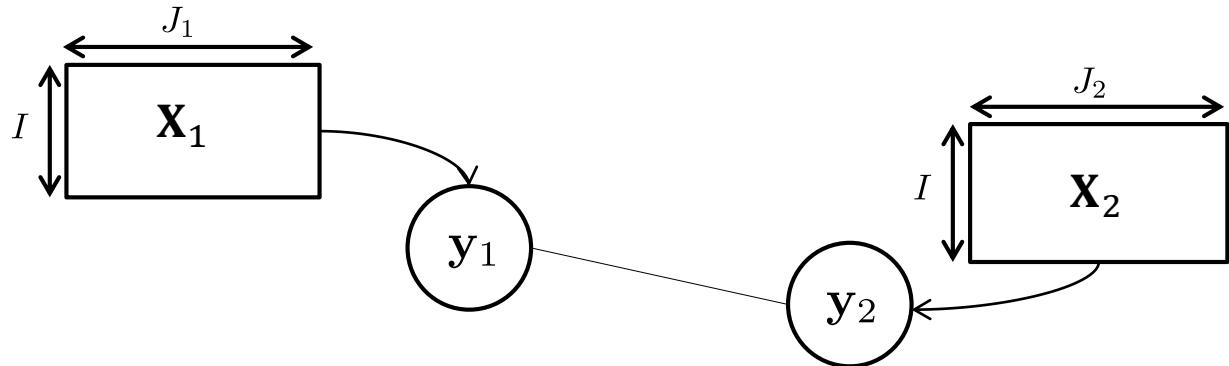
$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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

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



Regularized Generalized Canonical Correlation Analysis (RGCCA)

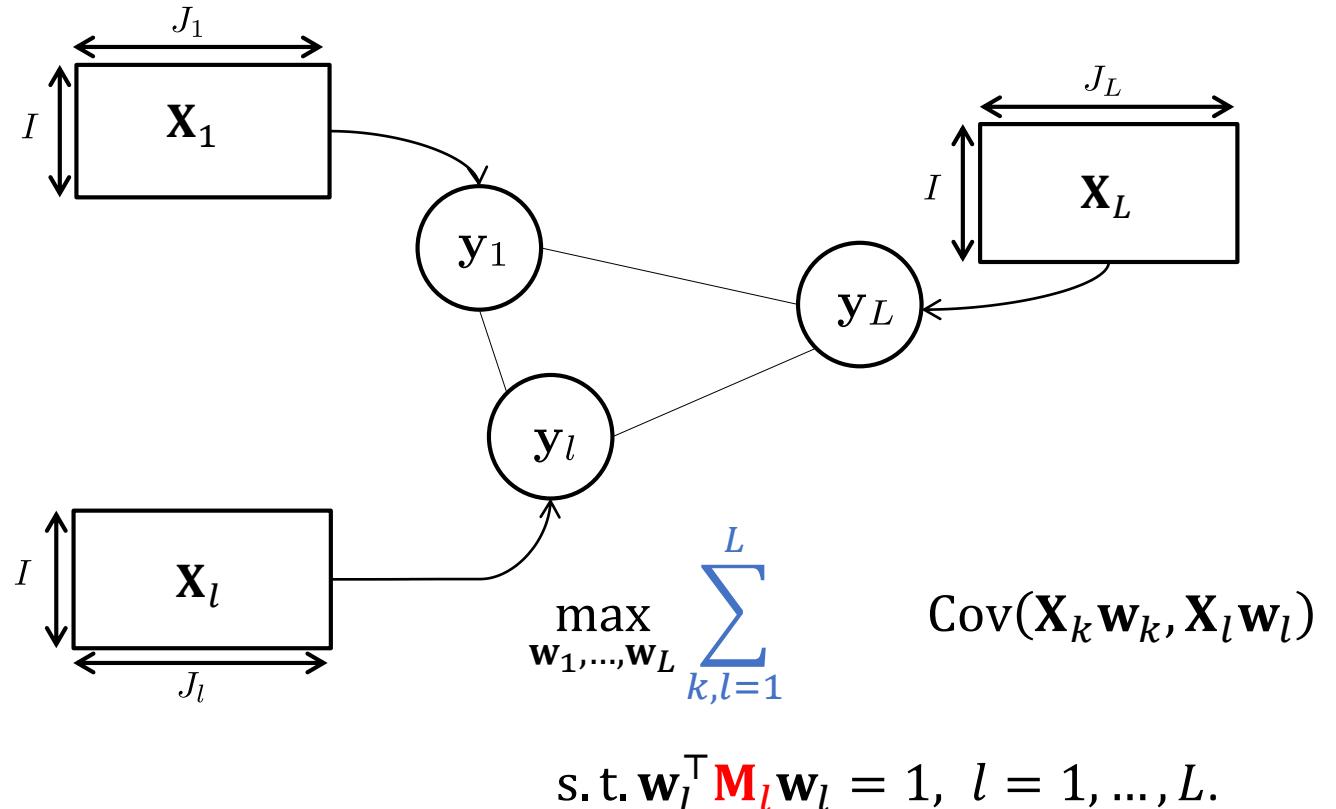


$$\max_{\mathbf{w}_1, \mathbf{w}_2} \text{Cov}(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2)$$

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

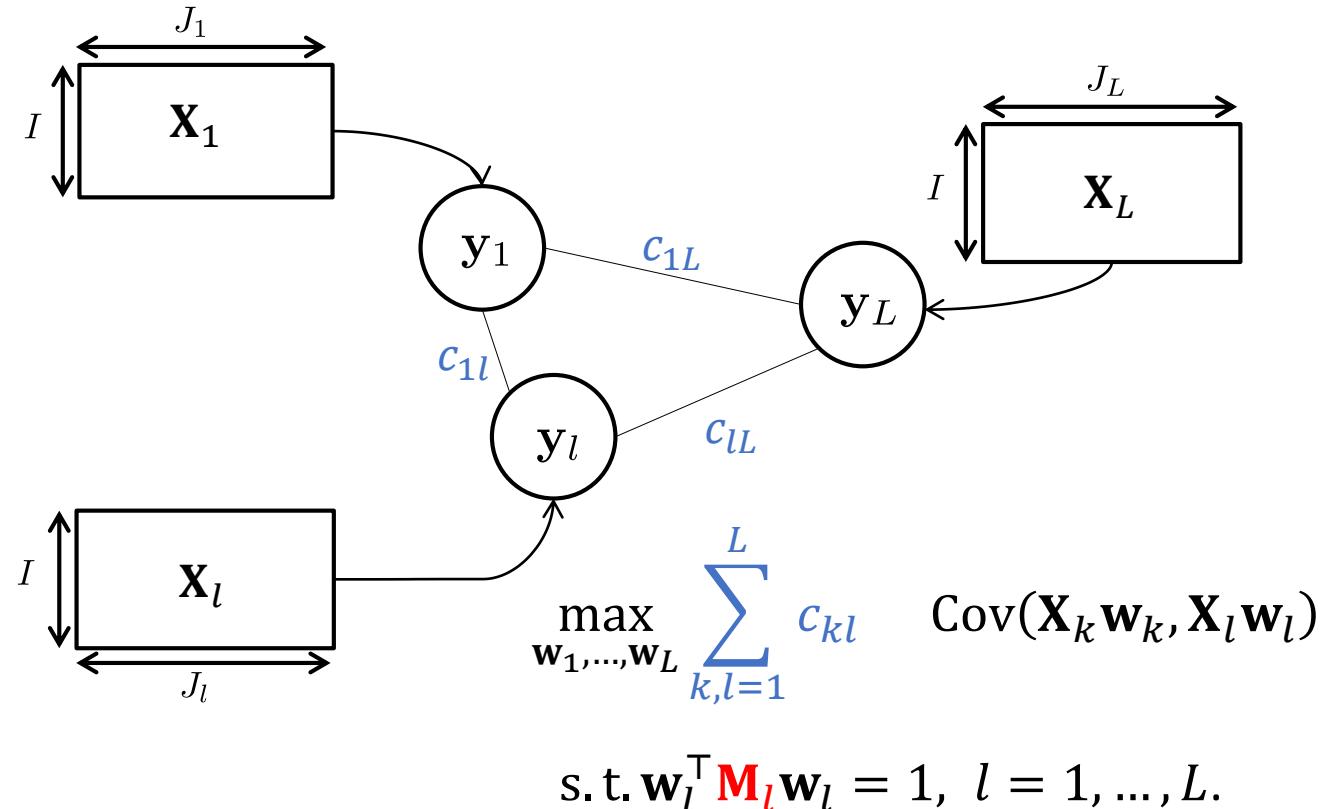


Regularized Generalized Canonical Correlation Analysis (RGCCA)



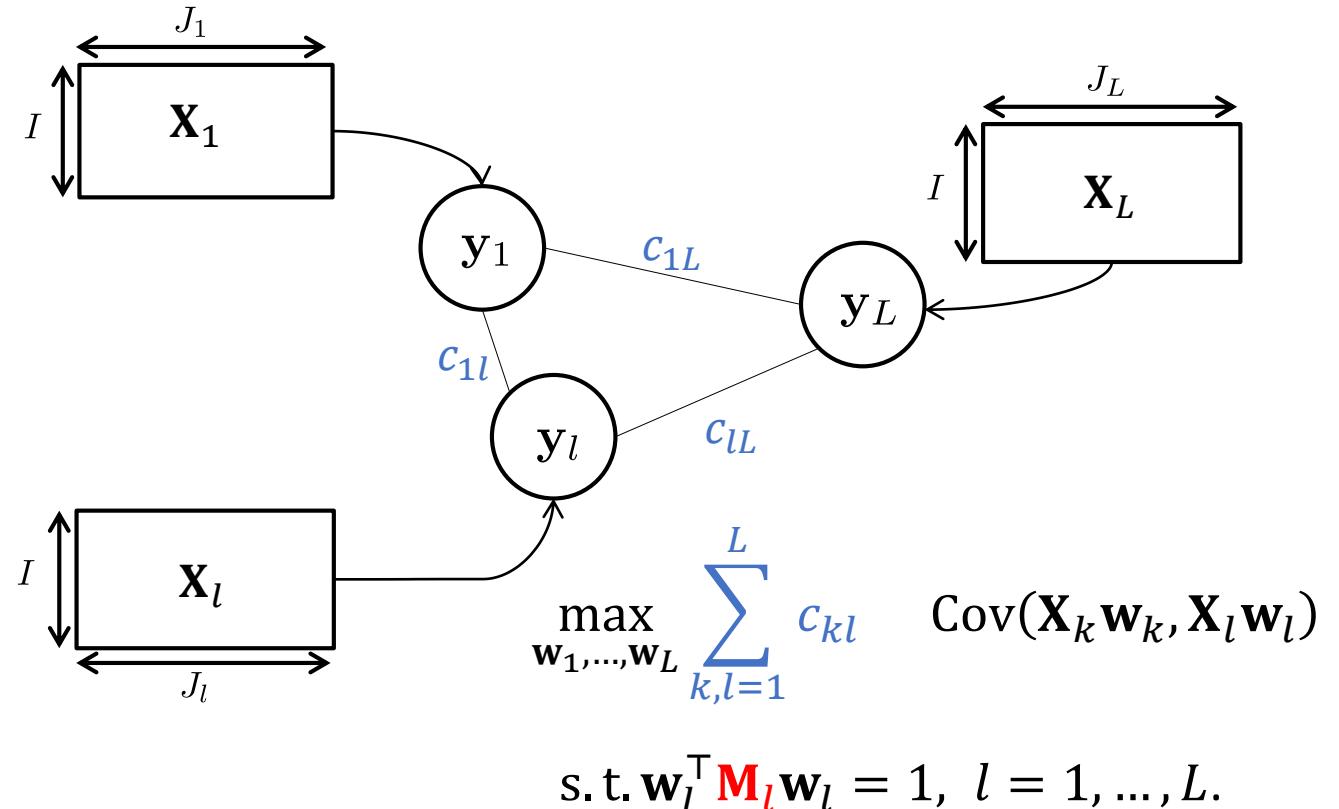


Regularized Generalized Canonical Correlation Analysis (RGCCA)





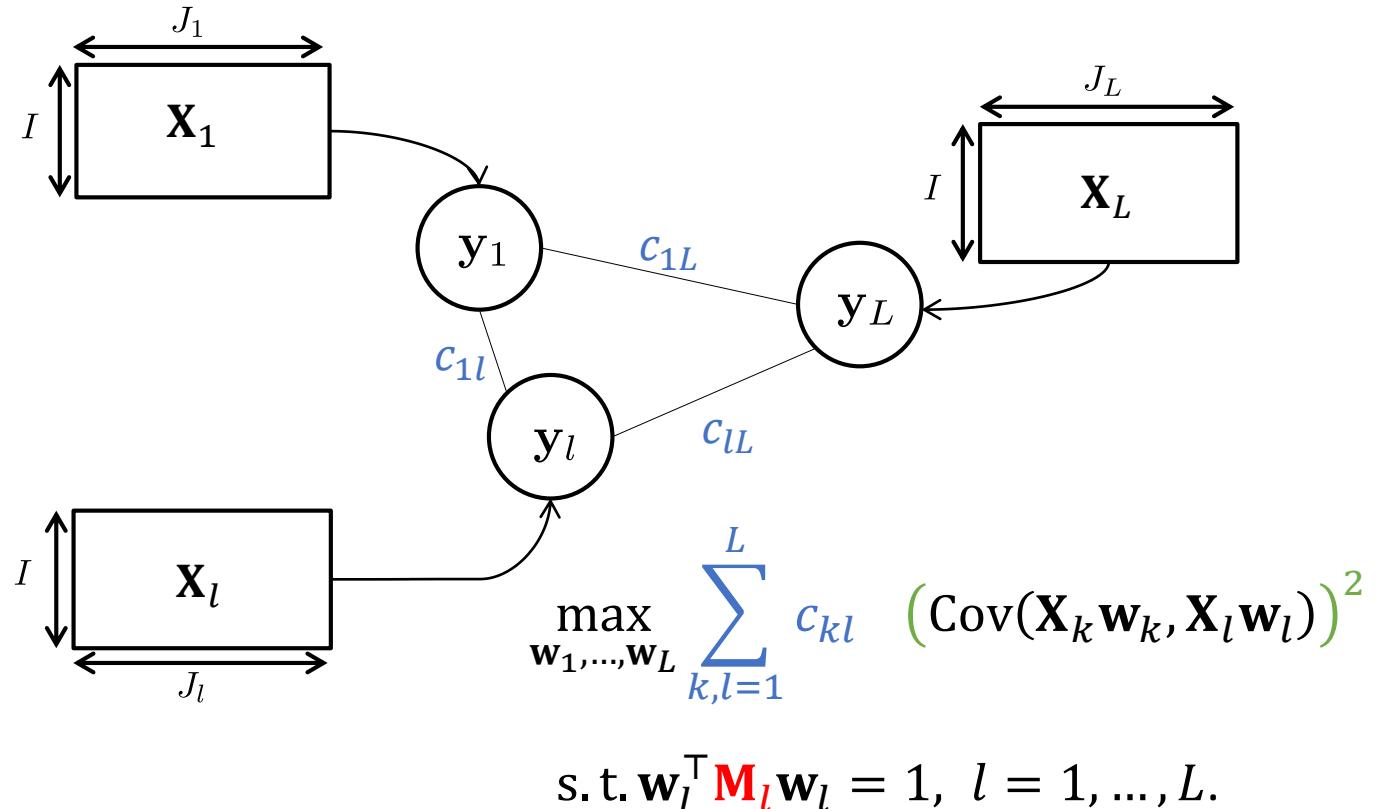
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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



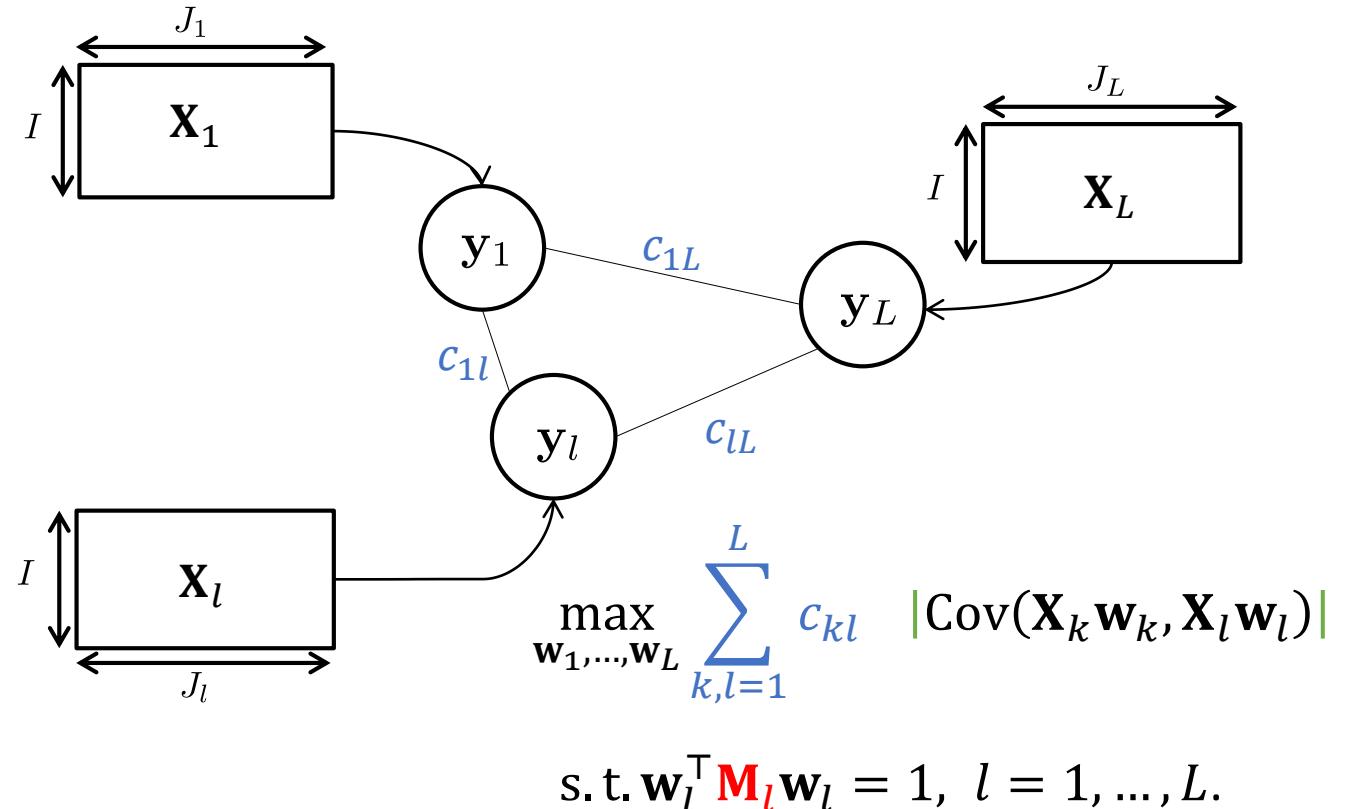
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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



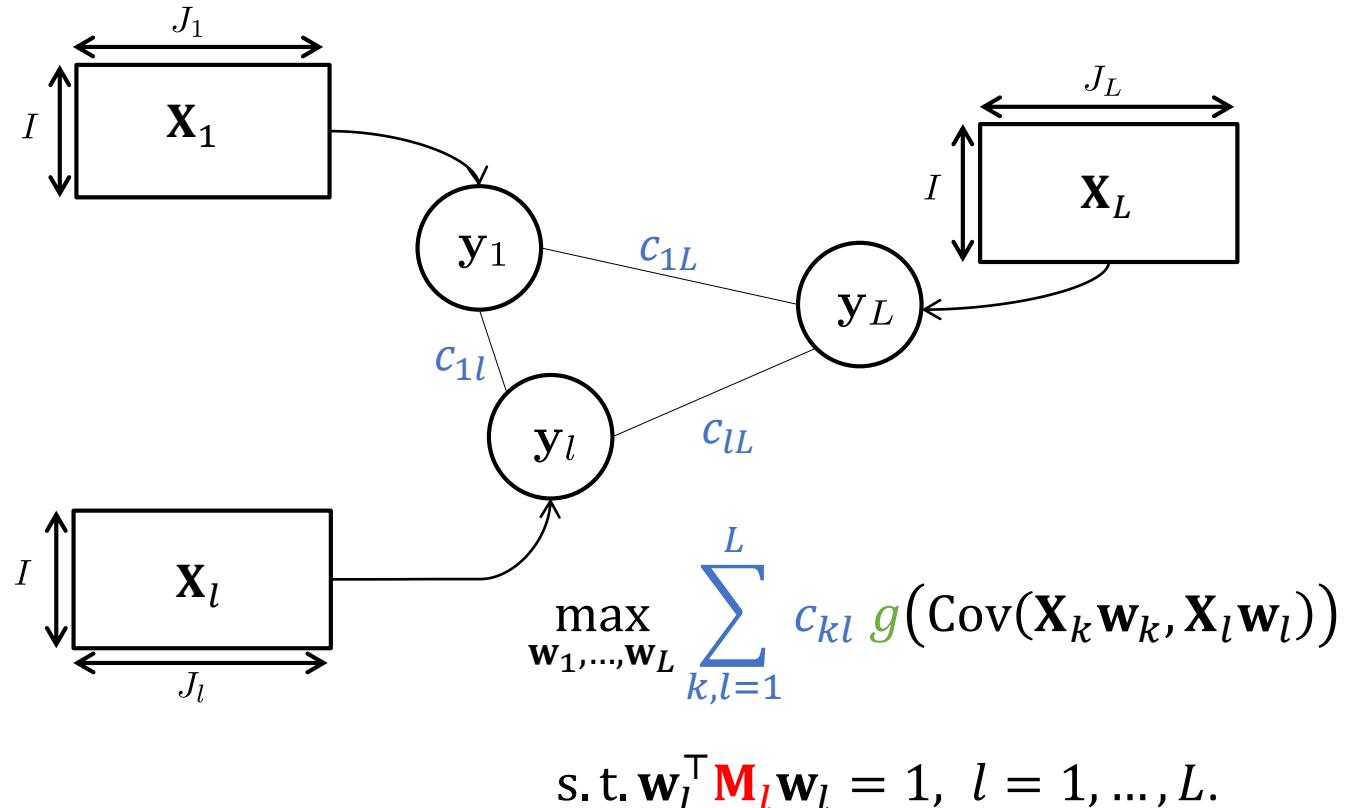
Regularized Generalized Canonical Correlation Analysis (RGCCA)



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

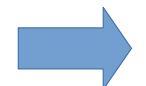
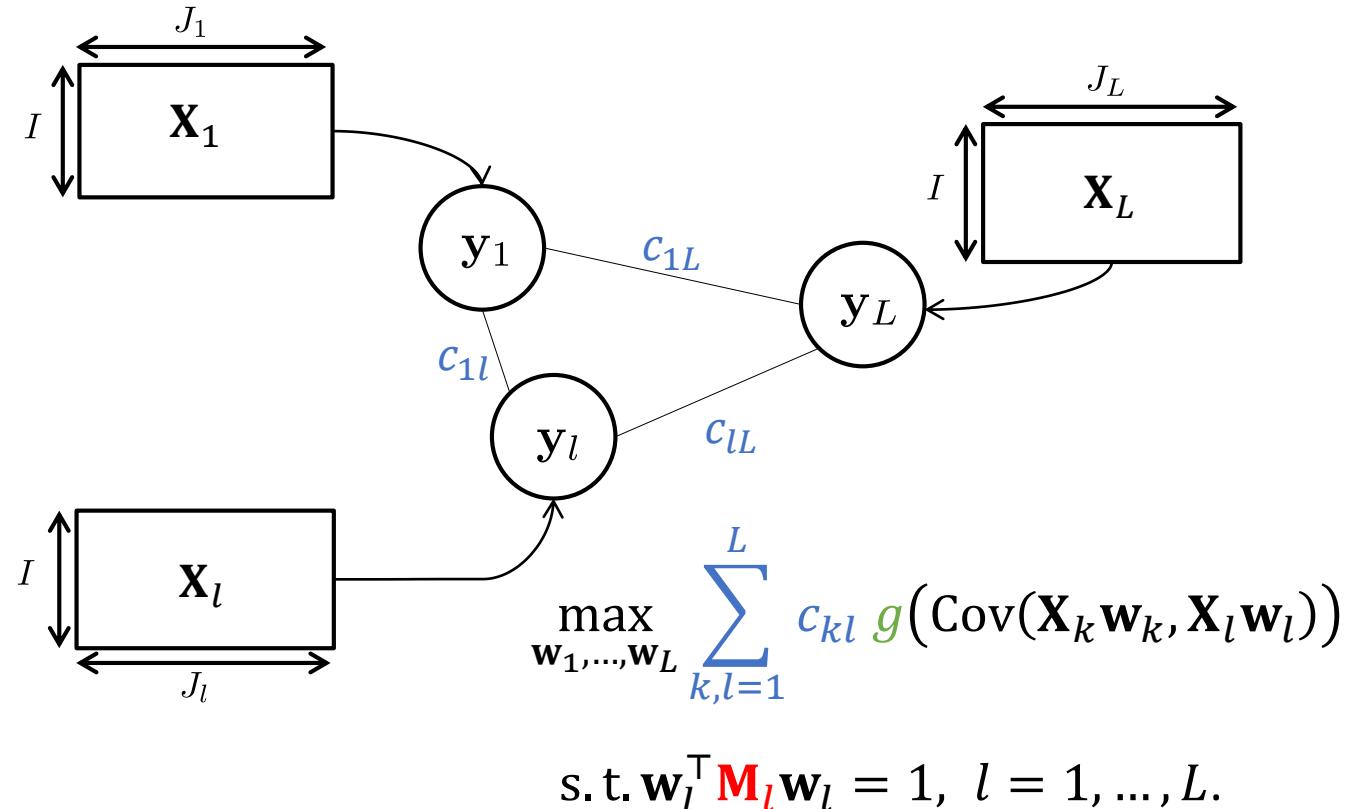


Regularized Generalized Canonical Correlation Analysis (RGCCA)





Regularized Generalized Canonical Correlation Analysis (RGCCA)



with g a continuous, convex and derivable function.



Summary of RGCCA

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

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

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

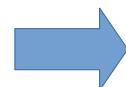


Summary of RGCCA

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

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

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



With “ g ” a continuous, convex and derivable function.



Summary of RGCCA

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

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

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

- 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 :

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

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

- 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.



Summary of RGCCA

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

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

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

- 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:



Summary of RGCCA

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

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

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

- 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:

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



Summary of RGCCA

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

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

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

- 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 :

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

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

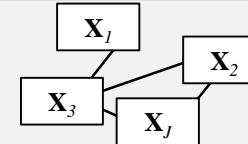
- With “ g ” a continuous, convex and derivable function.
- $c_{lk} = 1$ for two connected blocks and 0 otherwise.
- Where \mathbf{M}_l is any $J_l \times J_l$ positive definite matrix.

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

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

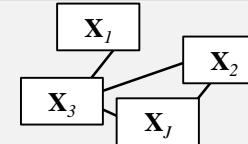
Overview of the Multi-Block litterature

BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$	
SUMCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SSQCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SABSCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k) $



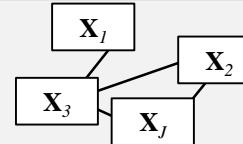
Overview of the Multi-Block litterature

BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$	
SUMCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SSQCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SABSCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k) $



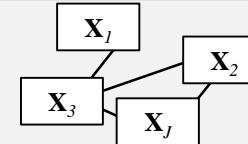
Overview of the Multi-Block litterature

BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$	
SUMCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SSQCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SABSCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k) $



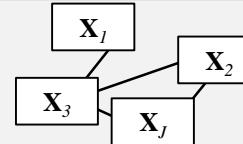
Overview of the Multi-Block litterature

BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$	
SUMCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SSQCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SABSCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k) $



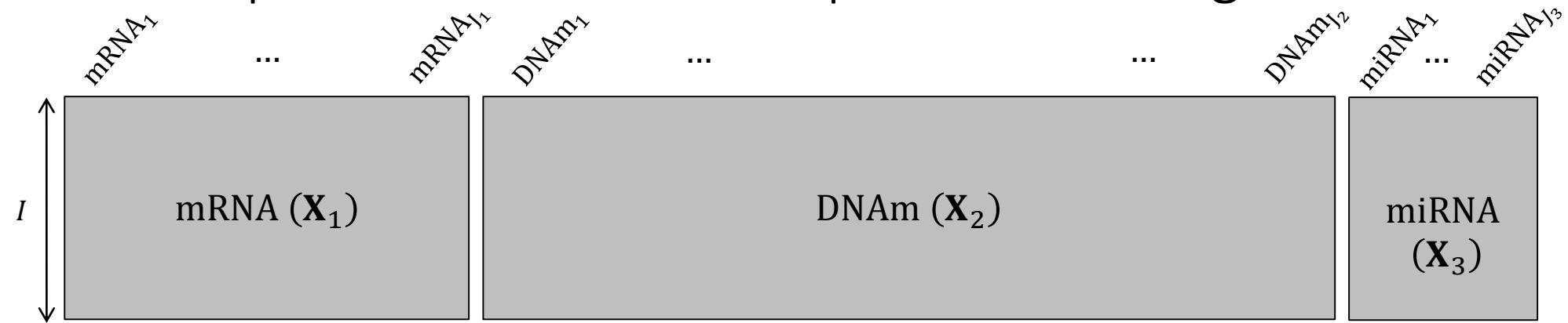
Overview of the Multi-Block litterature

BLOCKS ARE PARTIALLY CONNECTED $c_{jk} = 1 \text{ if } \mathbf{X}_j \leftrightarrow \mathbf{X}_k, 0 \text{ otherwise}$	
SUMCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SSQCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}^2(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k)$
SABSCOR	$\max_{\text{var}(\mathbf{X}_j \mathbf{w}_j) = 1} \sum_{j,k} c_{jk} \text{cov}(\mathbf{X}_j \mathbf{w}_j, \mathbf{X}_k \mathbf{w}_k) $

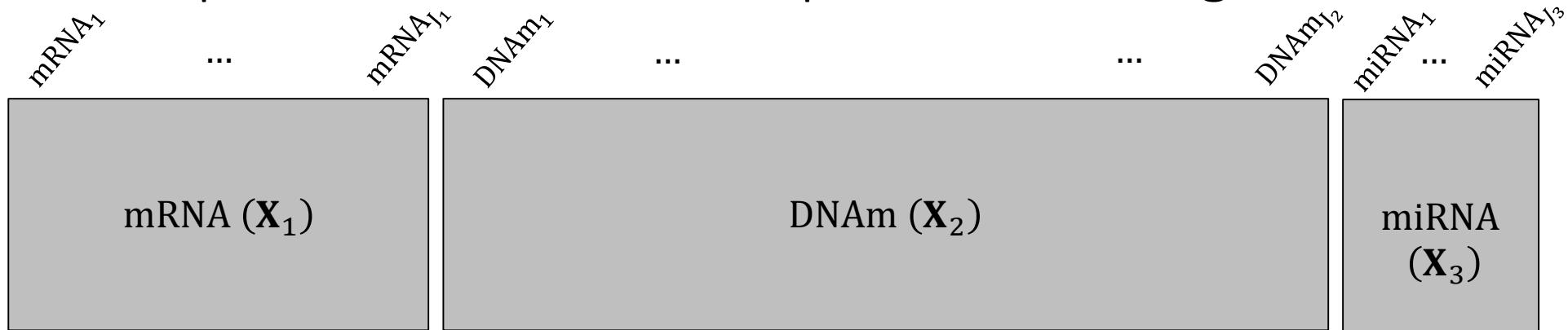


Courtesy to Arthur Tenenhaus.

Tune parameters in an unsupervised setting



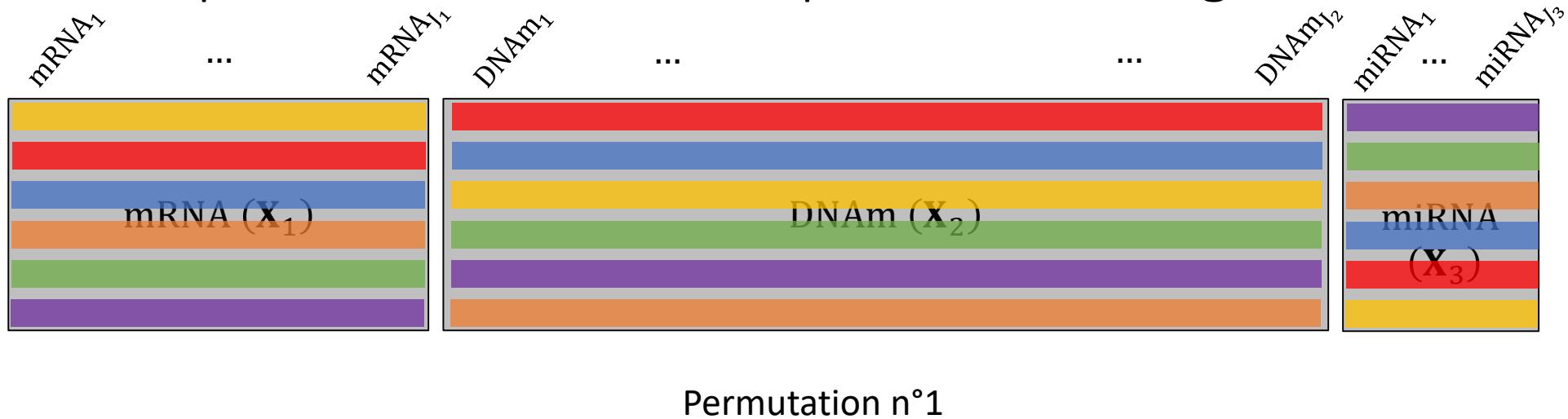
Tune parameters in an unsupervised setting



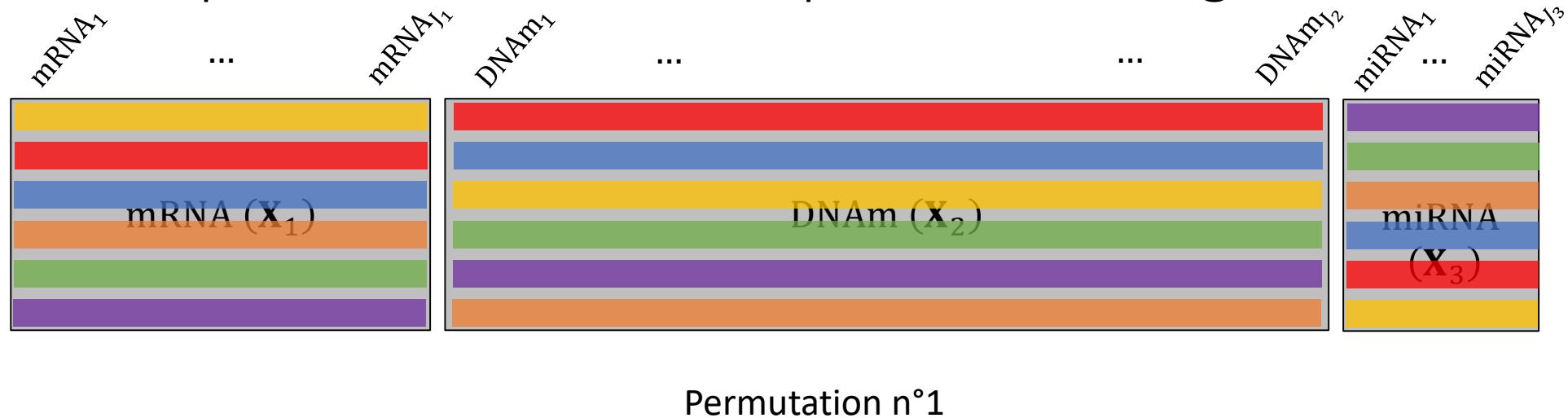
Tune parameters in an unsupervised setting



Tune parameters in an unsupervised setting

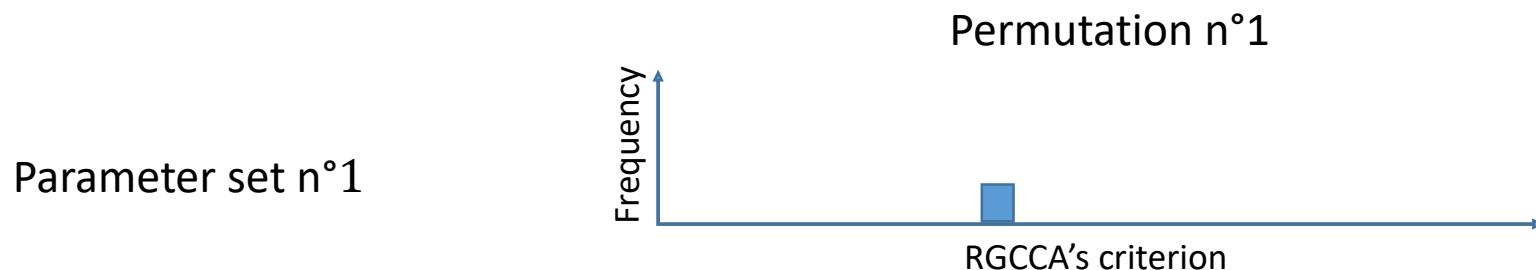


Tune parameters in an unsupervised setting



Parameter set n°1

Tune parameters in an unsupervised setting

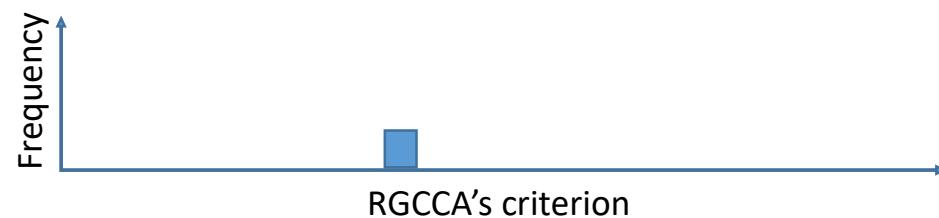


Tune parameters in an unsupervised setting



Permutation n°1

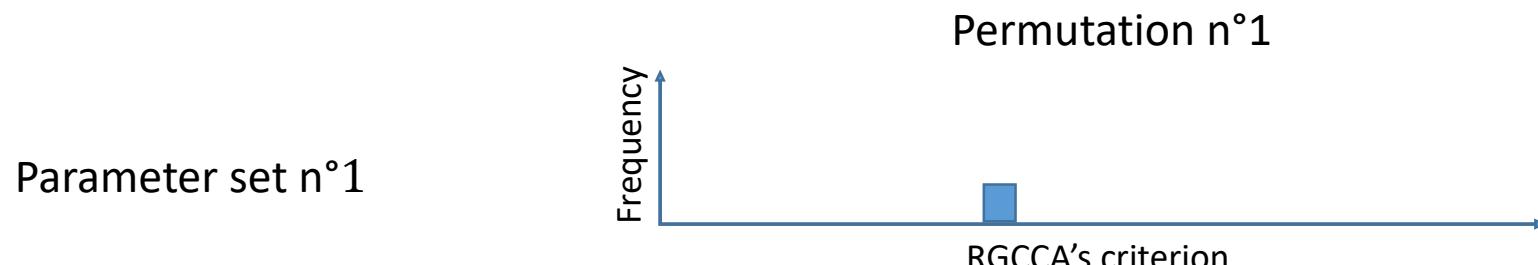
Parameter set n°1



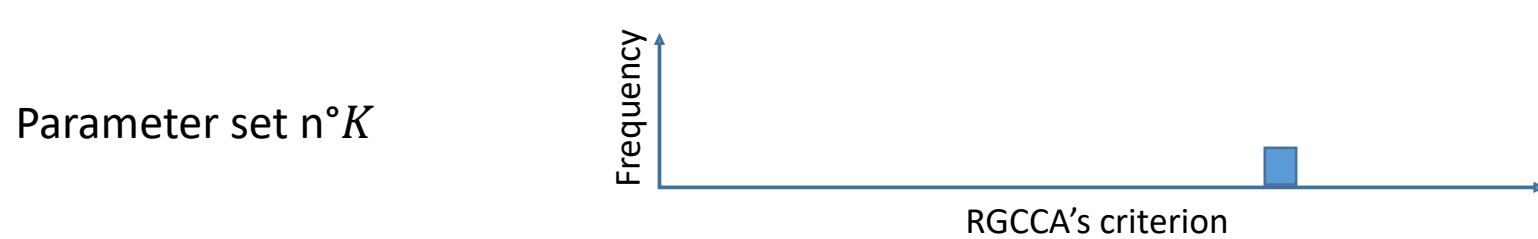
⋮

Parameter set n°K

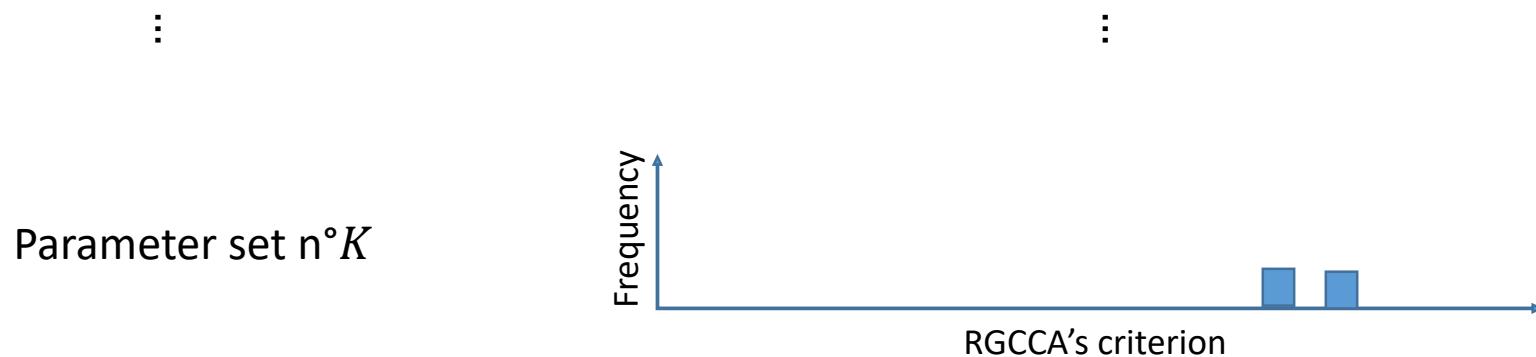
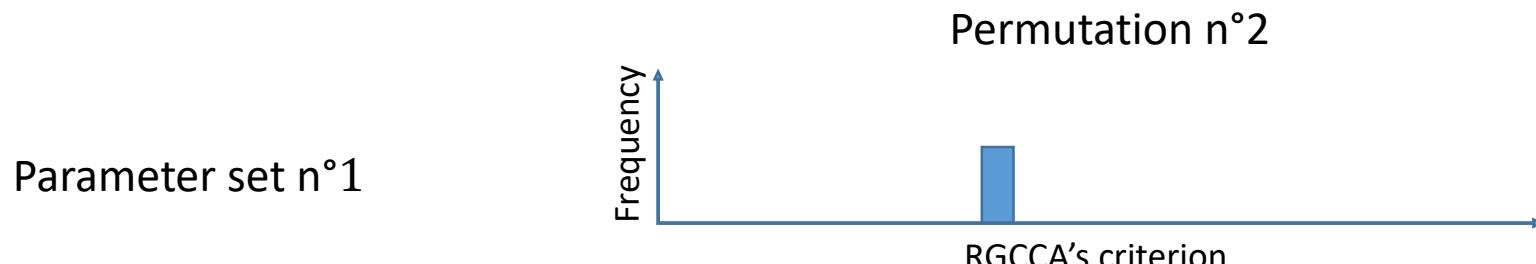
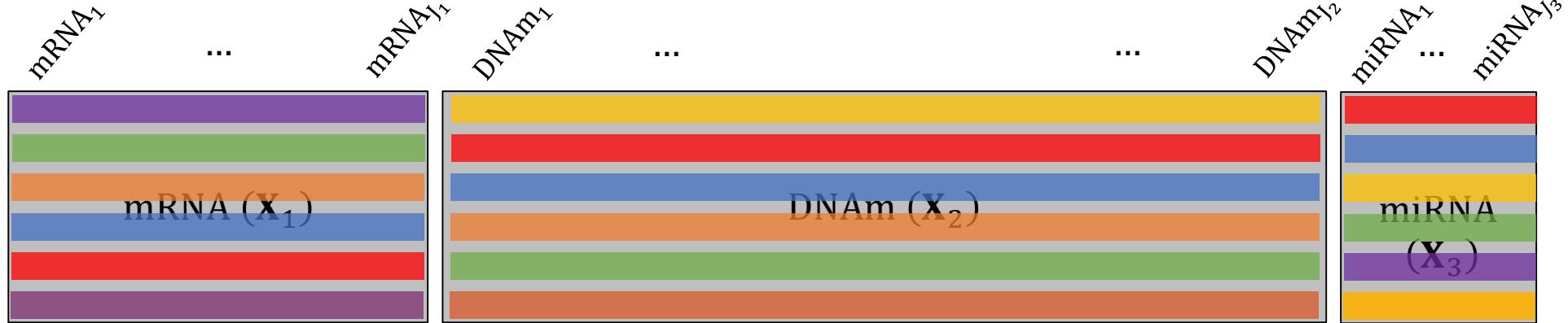
Tune parameters in an unsupervised setting



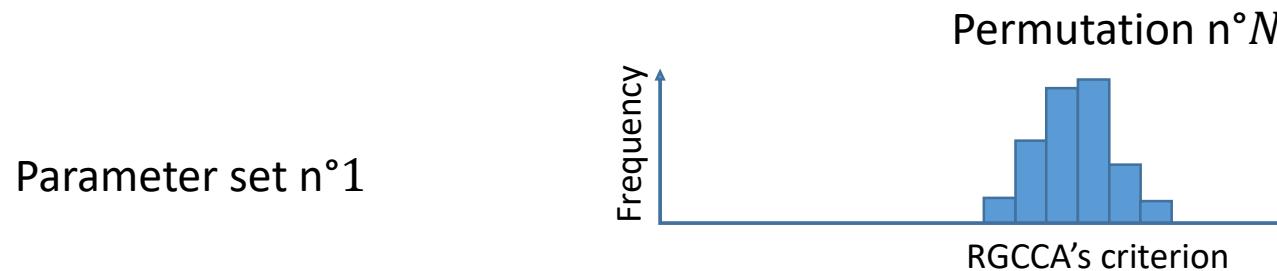
⋮



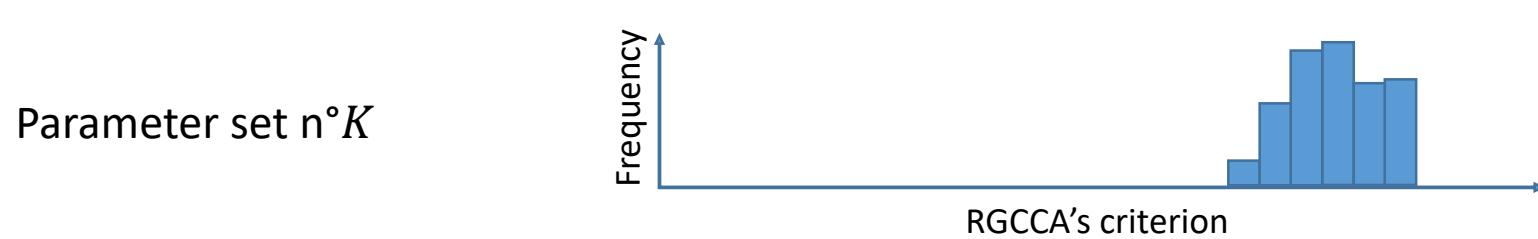
Tune parameters in an unsupervised setting



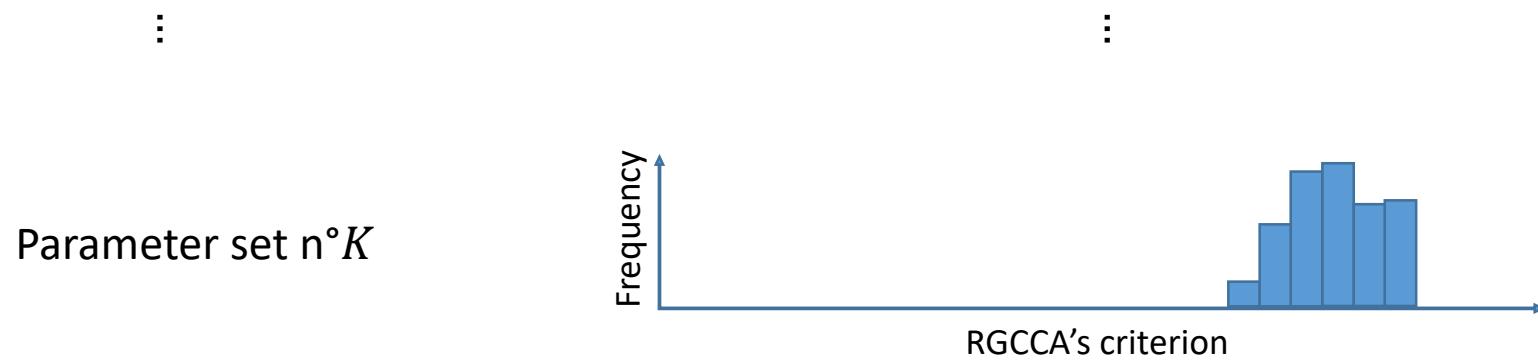
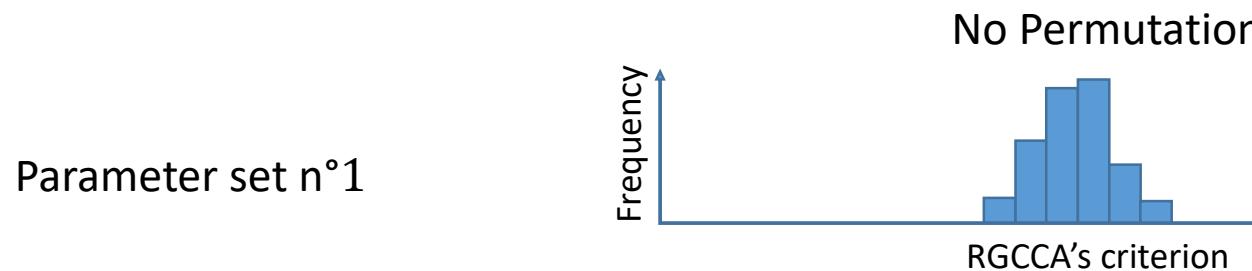
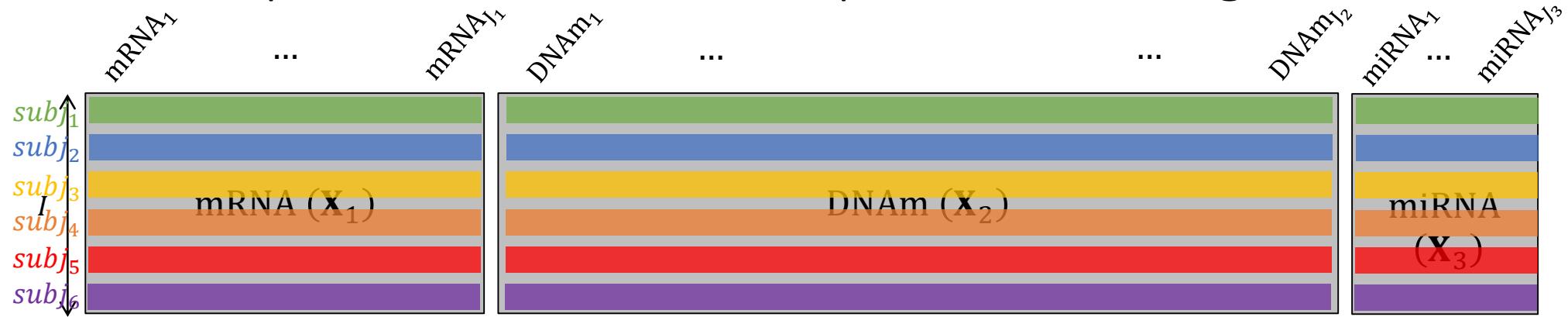
Tune parameters in an unsupervised setting



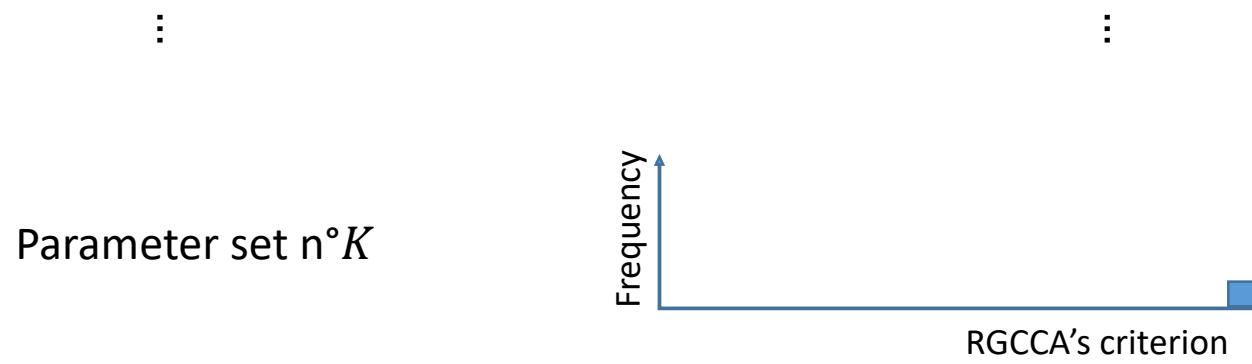
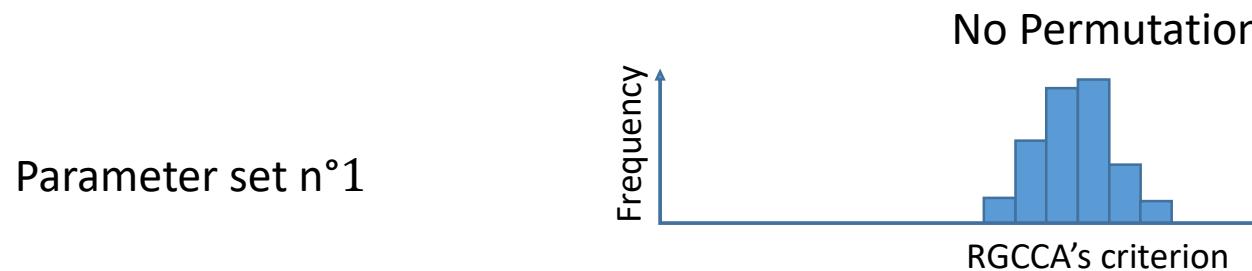
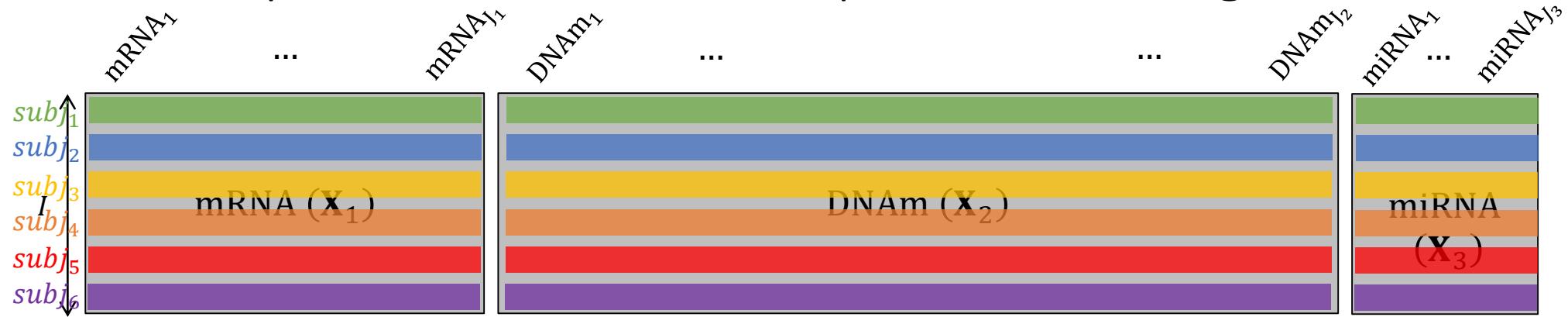
⋮



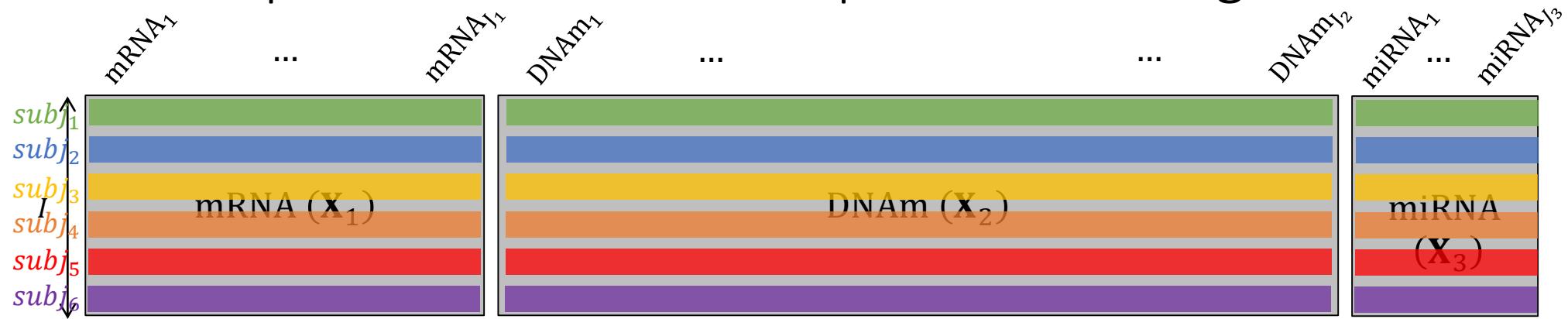
Tune parameters in an unsupervised setting



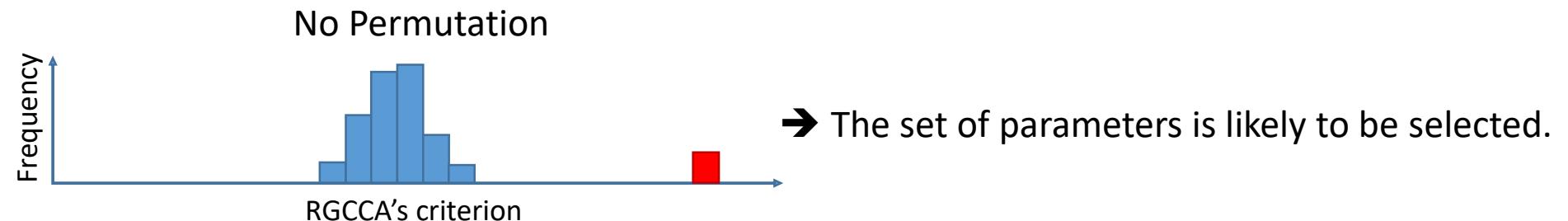
Tune parameters in an unsupervised setting



Tune parameters in an unsupervised setting



Parameter set n°1



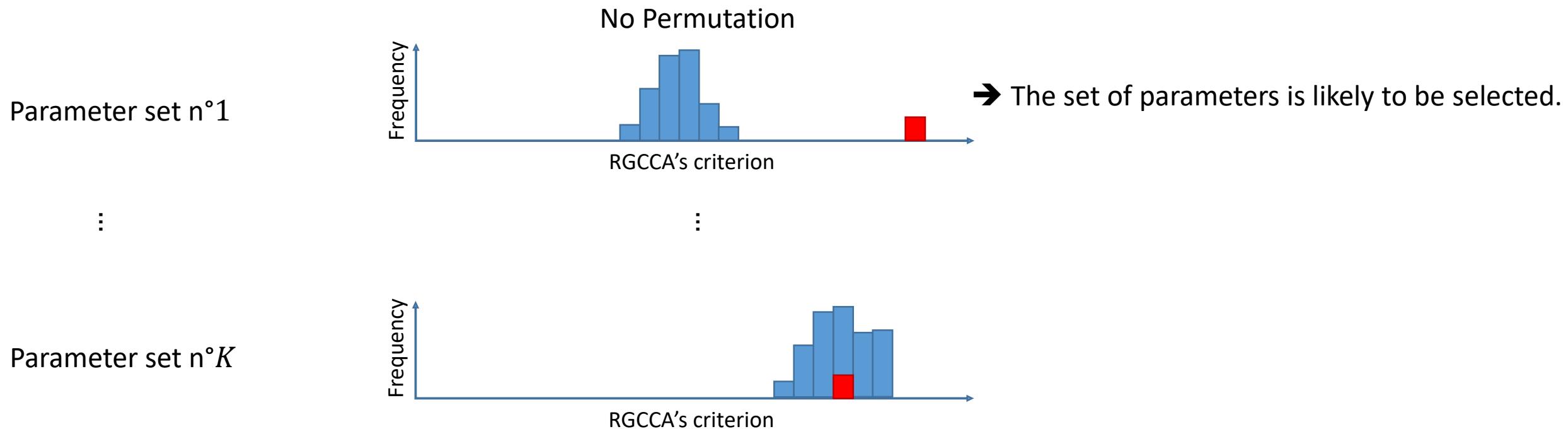
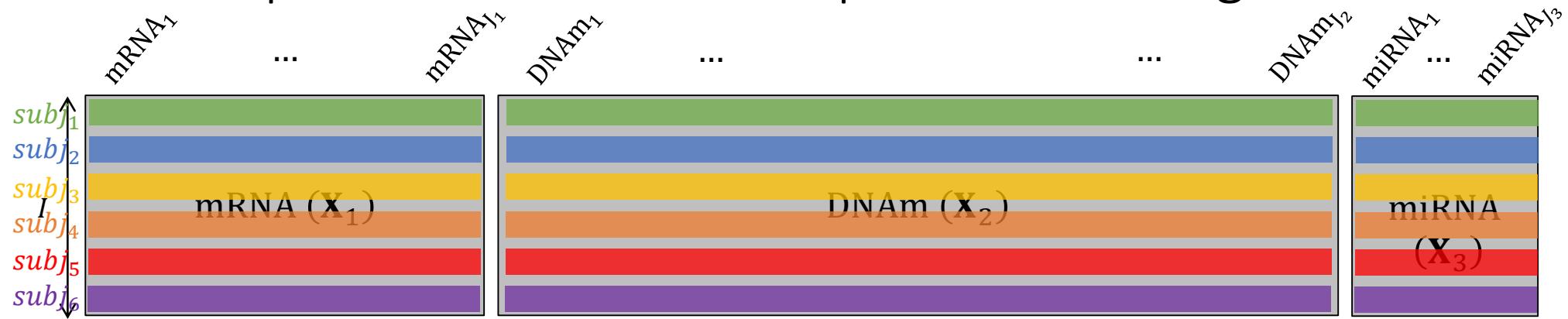
⋮

Parameter set n°K

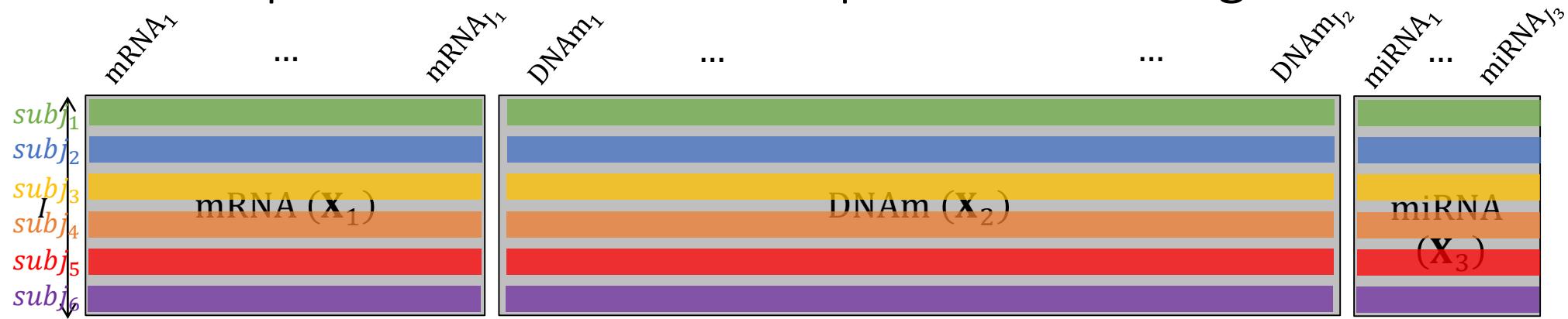


→ The set of parameters is likely to be selected.

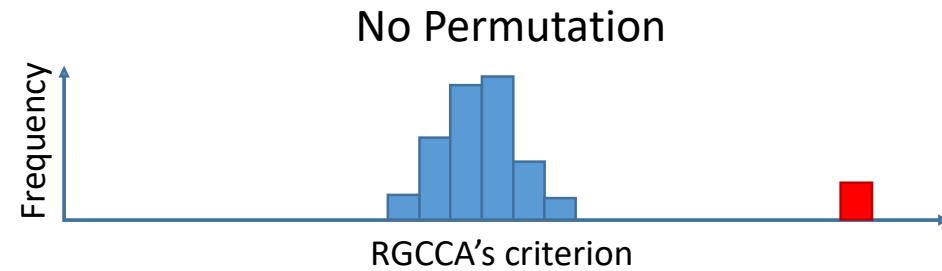
Tune parameters in an unsupervised setting



Tune parameters in an unsupervised setting



Parameter set n°1



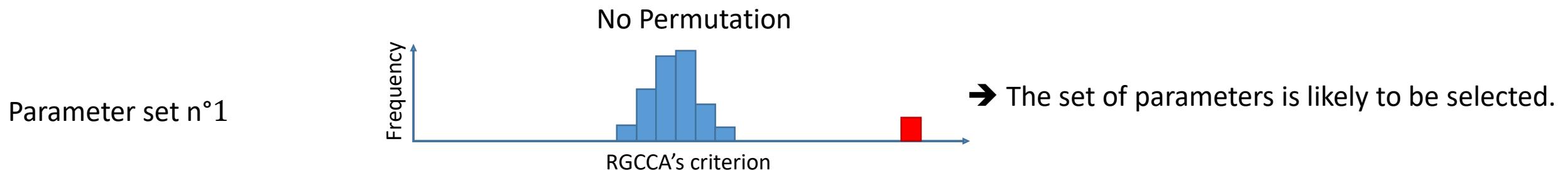
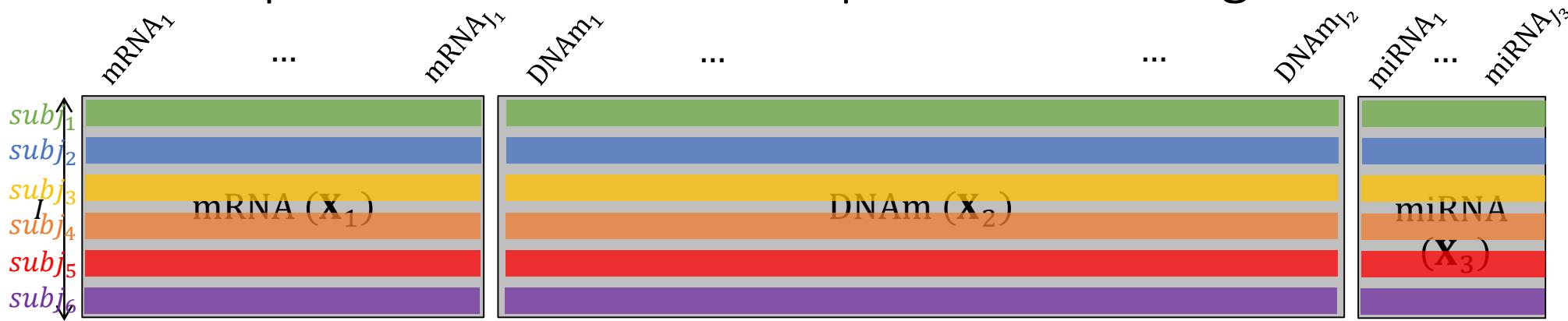
→ The set of parameters is likely to be selected.

Parameter set n°K

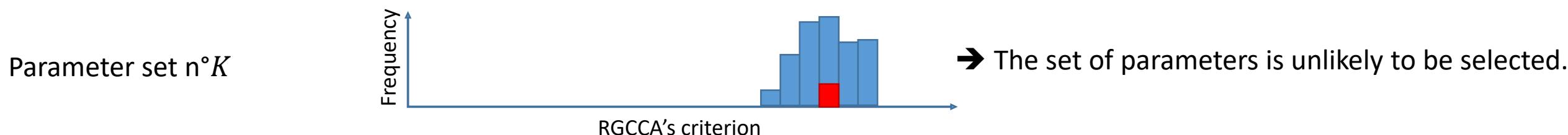


→ The set of parameters is unlikely to be selected.

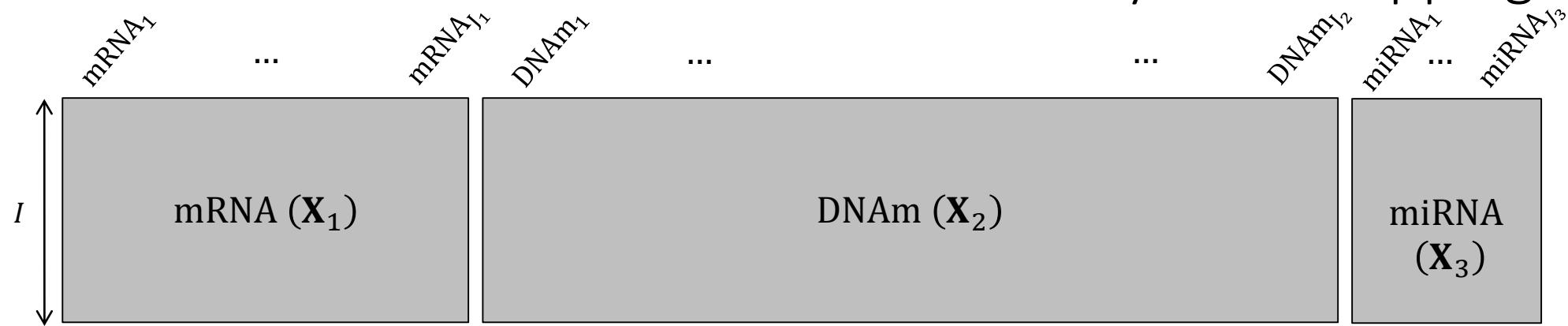
Tune parameters in an unsupervised setting



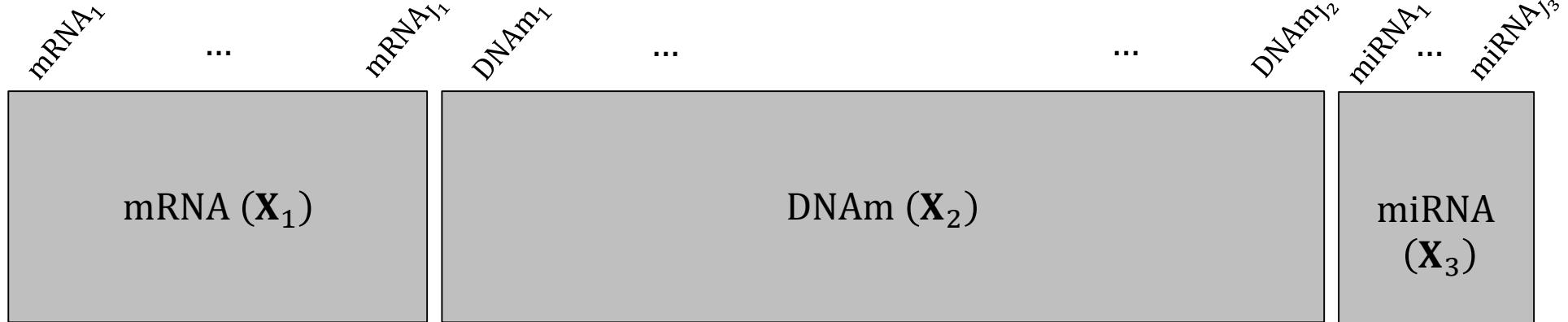
RGCCA choose the best set of parameters as the one with the highest value of $Z_k = \frac{(crit_{unperm} - \mu_{crit}^{perm})}{\sigma_{crit}^{perm}}$



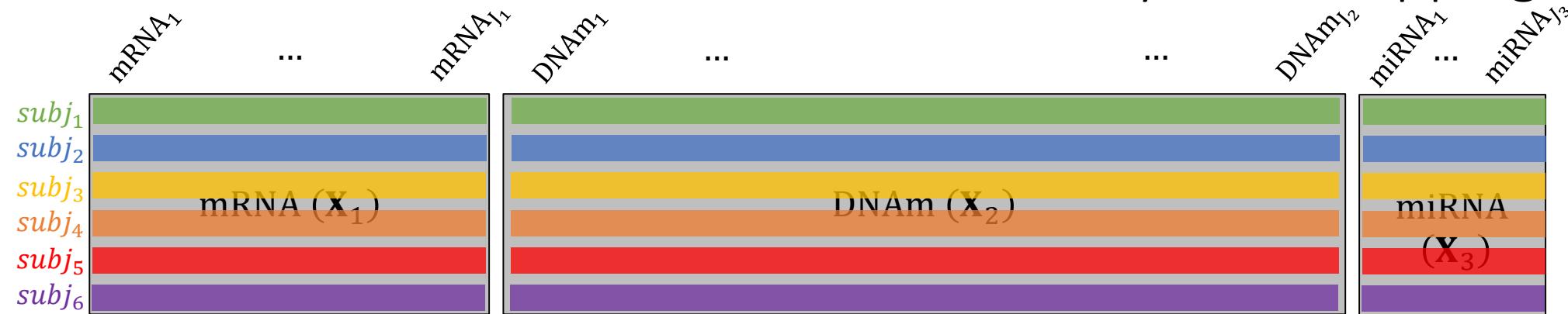
Evaluate the robustness of the model by bootstrapping.



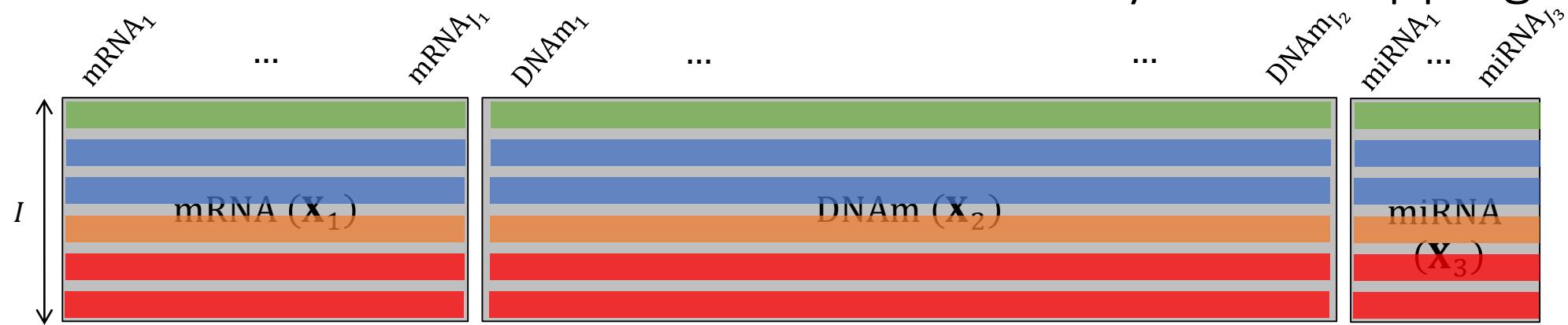
Evaluate the robustness of the model by bootstrapping.



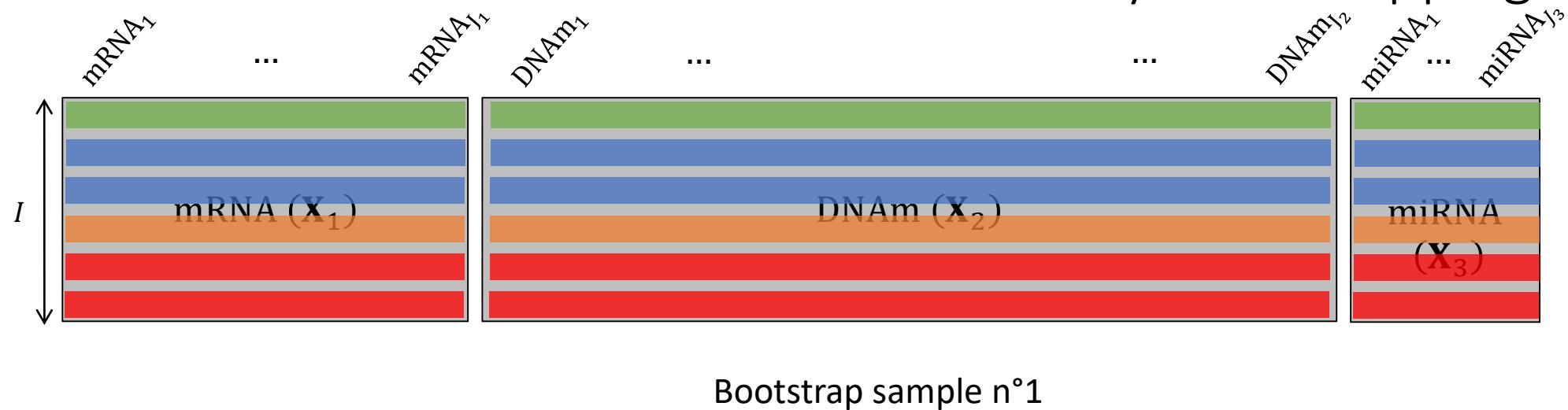
Evaluate the robustness of the model by bootstrapping.



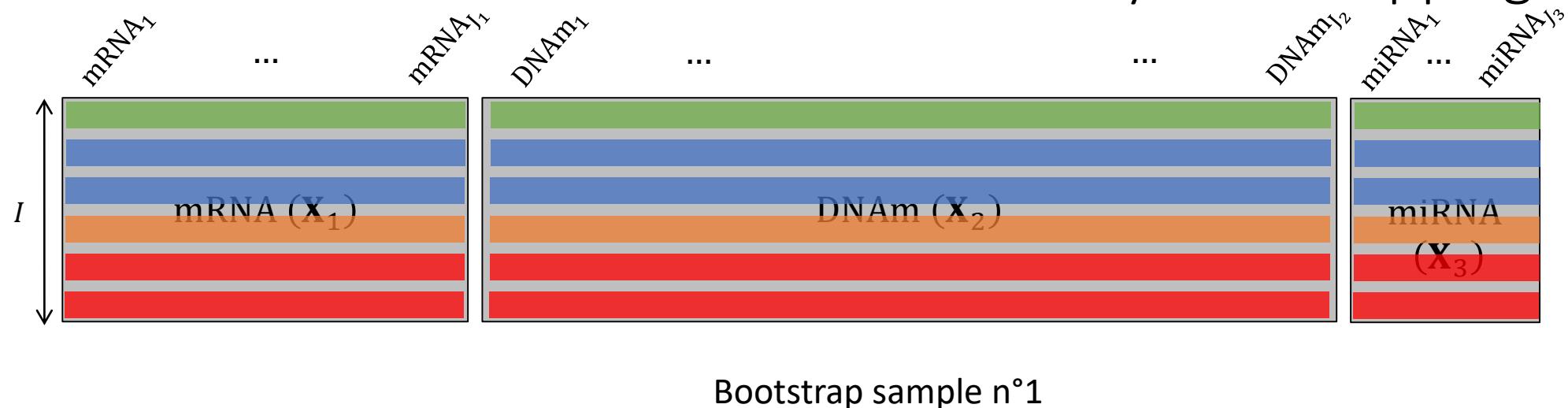
Evaluate the robustness of the model by bootstrapping.



Evaluate the robustness of the model by bootstrapping.

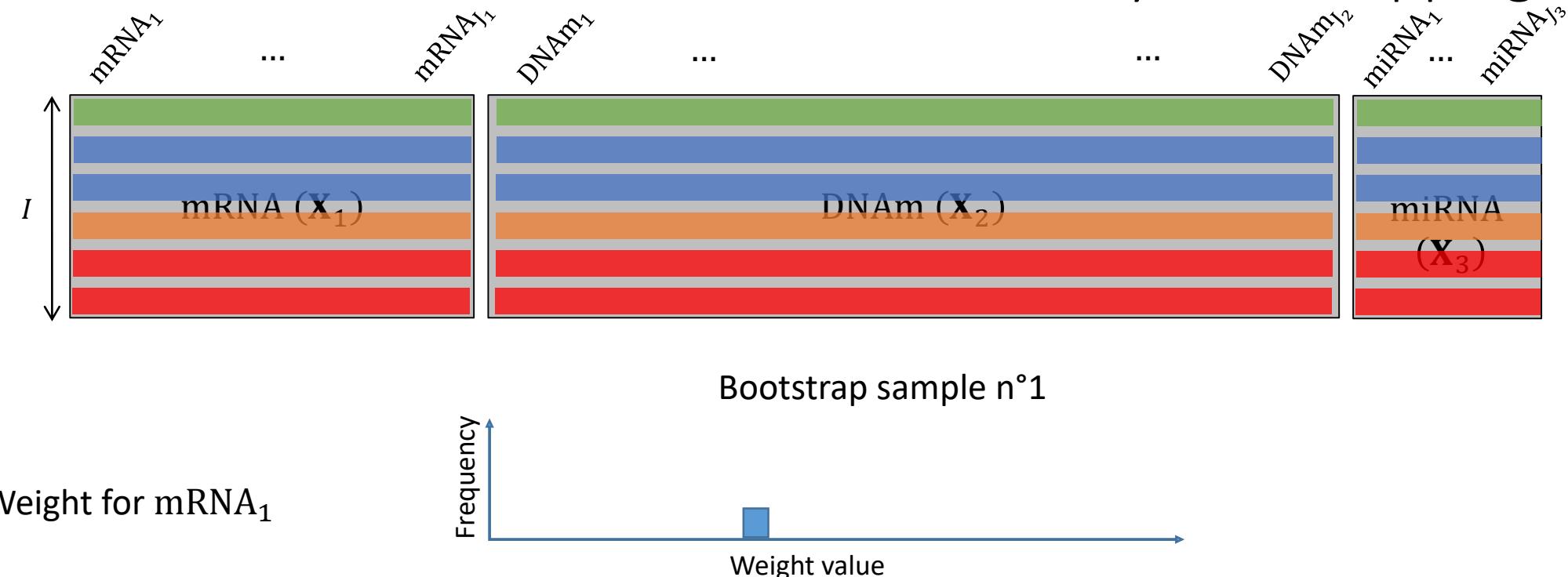


Evaluate the robustness of the model by bootstrapping.

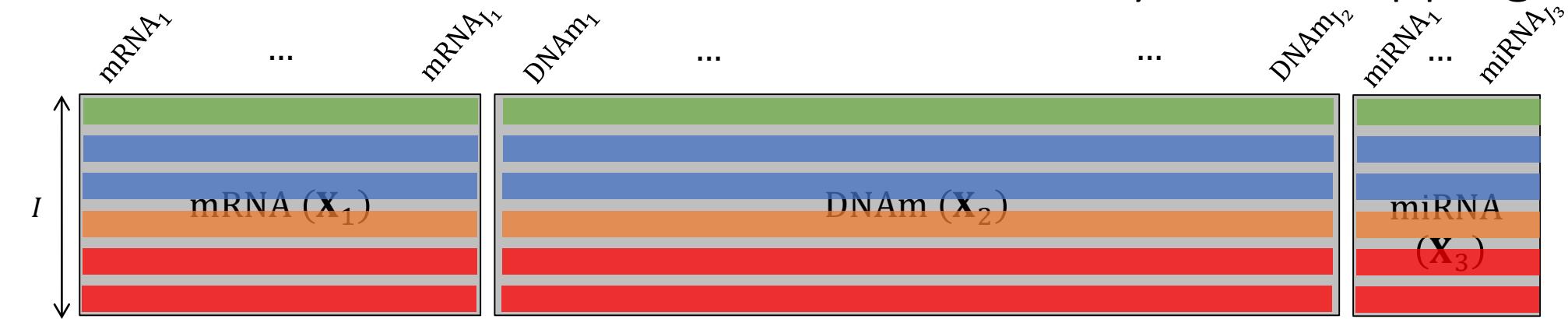


Weight for mRNA₁

Evaluate the robustness of the model by bootstrapping.

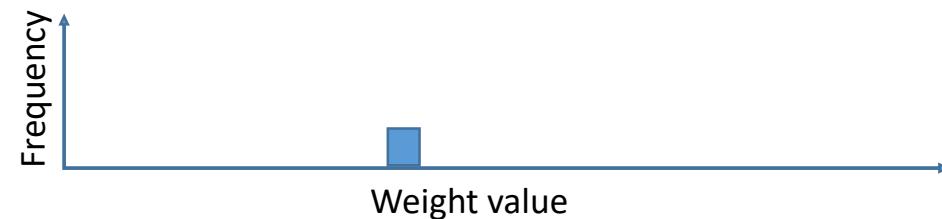


Evaluate the robustness of the model by bootstrapping.



Bootstrap sample n°1

Weight for $mRNA_1$



⋮

Weight for $miRNA_{J_3}$

Data ○○○

Unsupervised/1-block ○○○

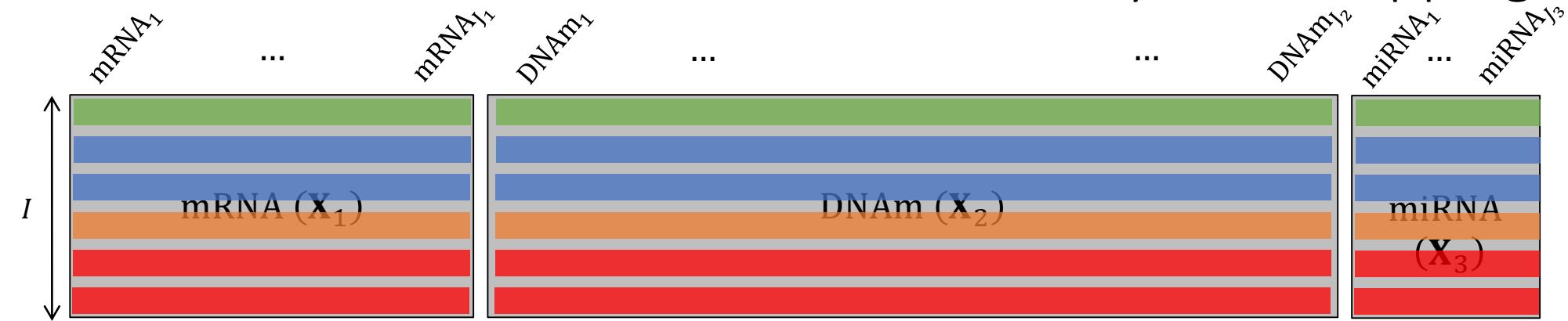
Unsup./2-blocks ○○○

Unsup./L-blocks ○○●

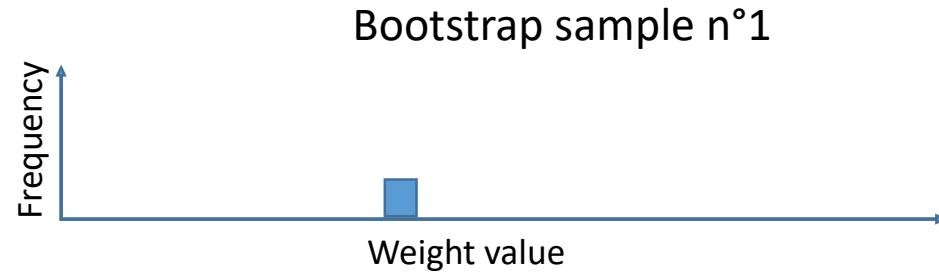
Supervised/L-blocks ○○

Further ○○○

Evaluate the robustness of the model by bootstrapping.

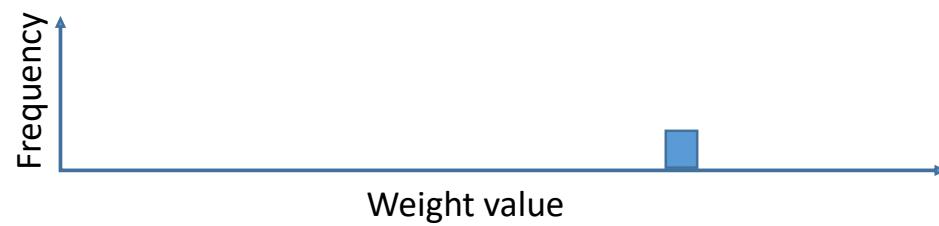


Weight for $mRNA_1$

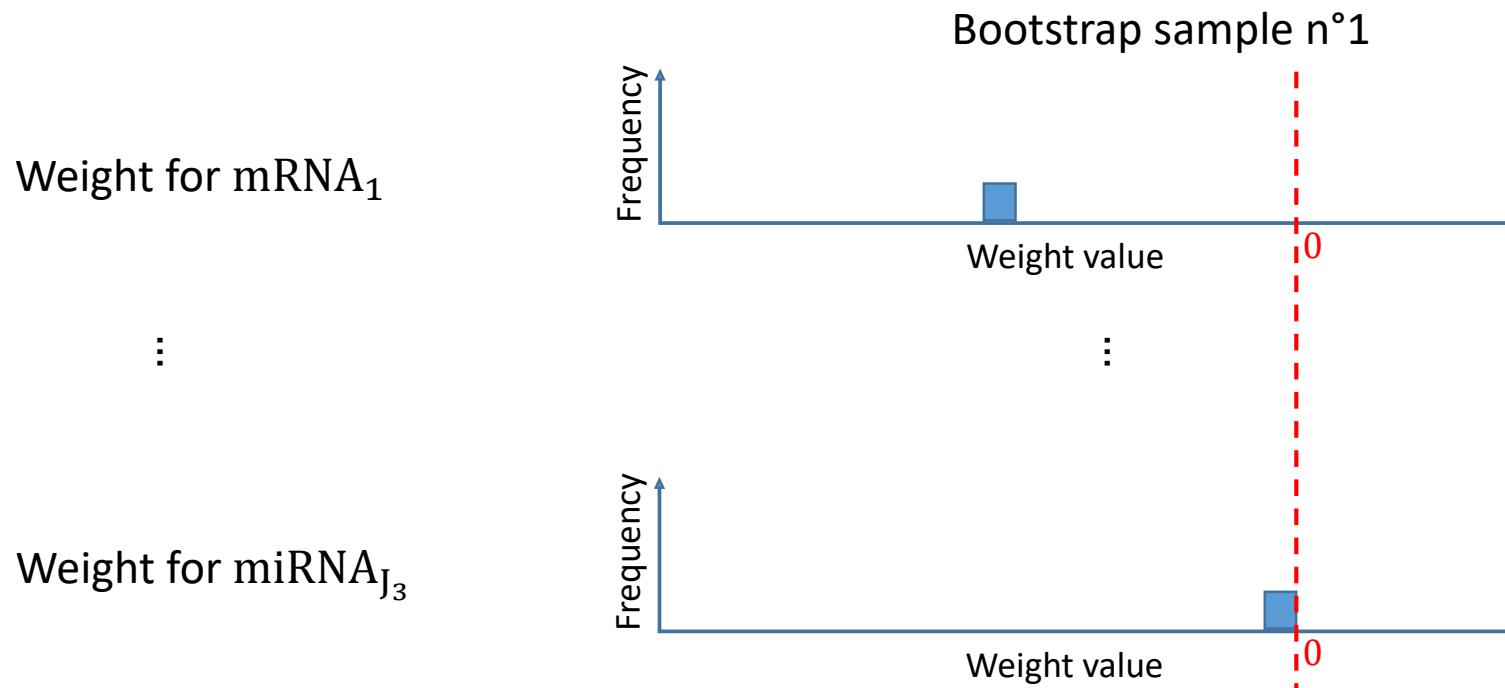
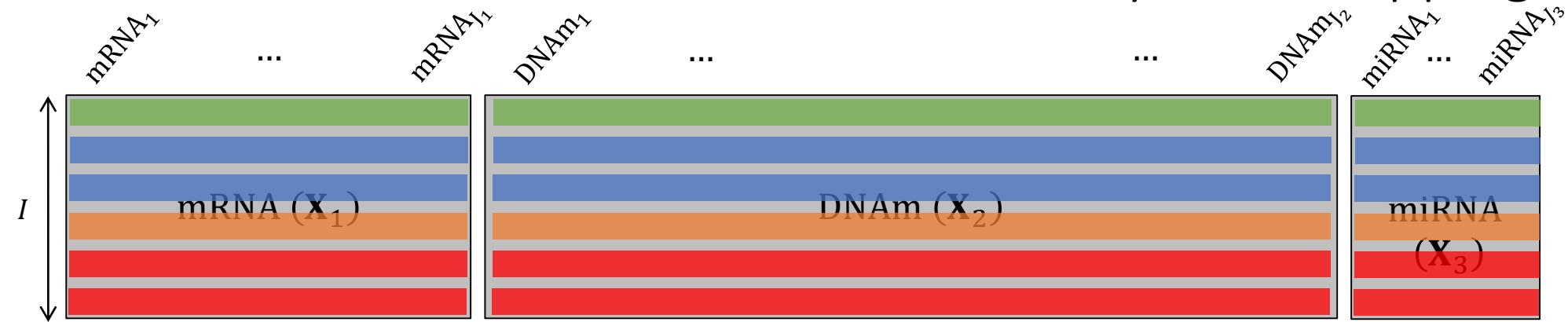


⋮

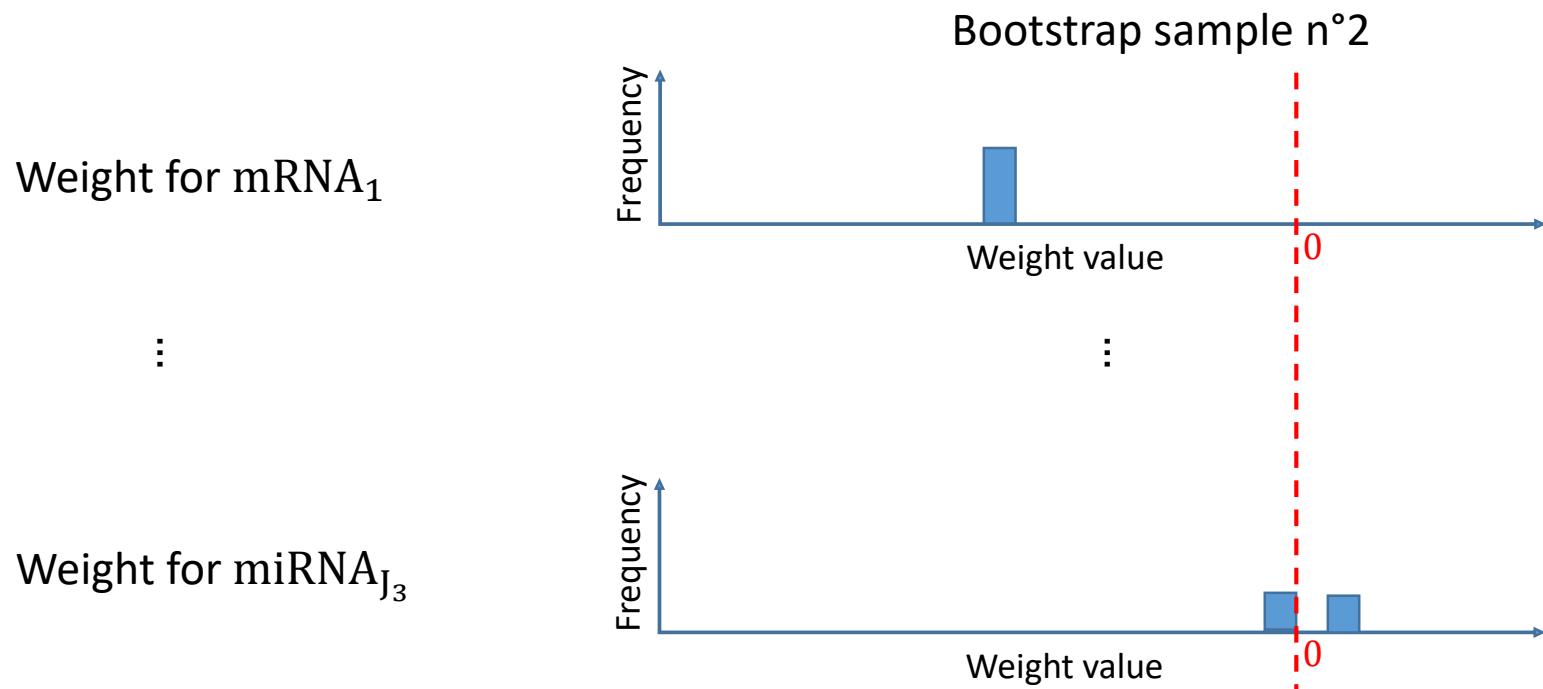
Weight for $miRNA_{J_3}$



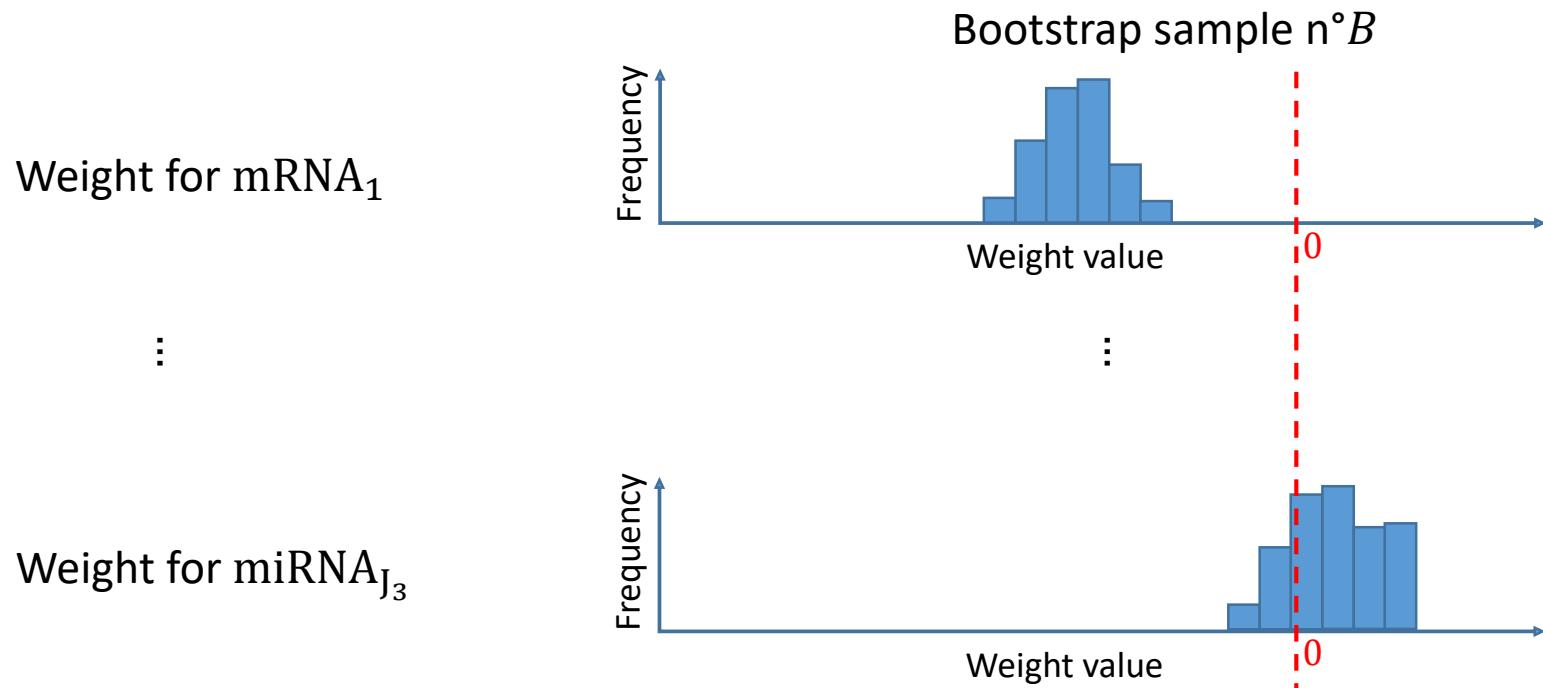
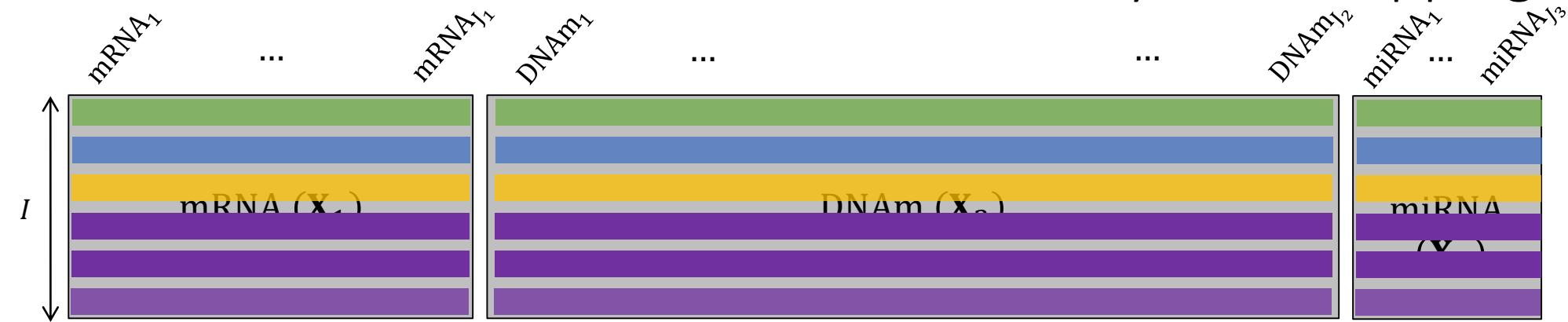
Evaluate the robustness of the model by bootstrapping.



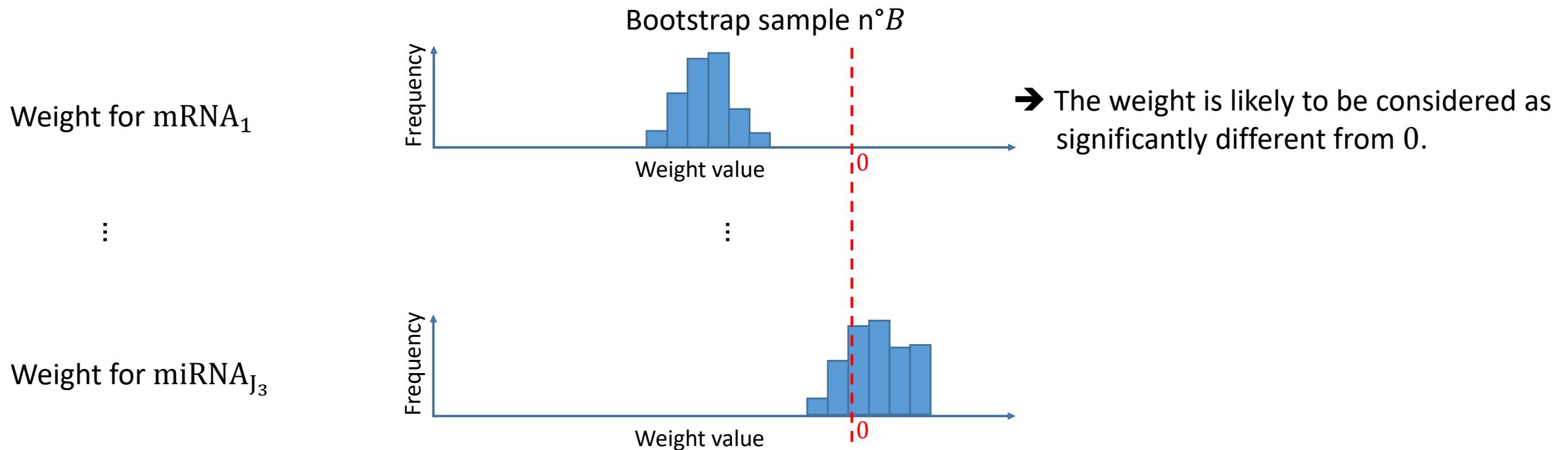
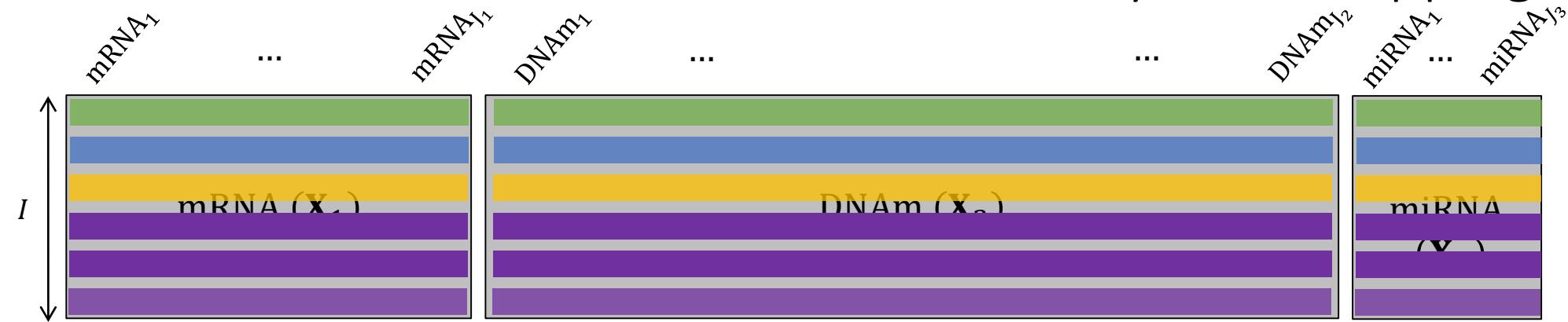
Evaluate the robustness of the model by bootstrapping.



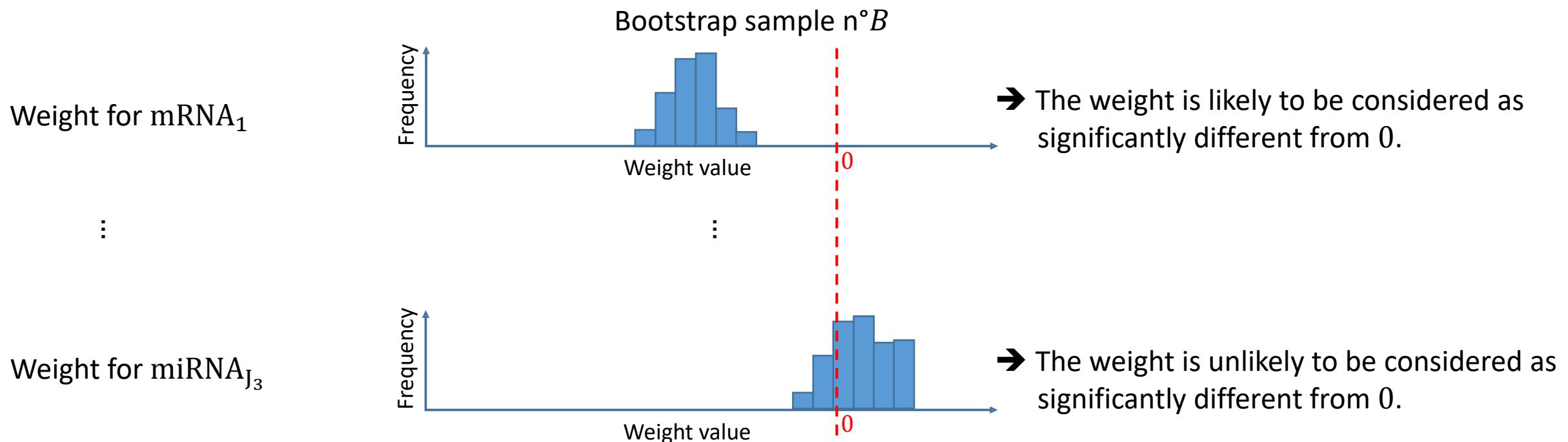
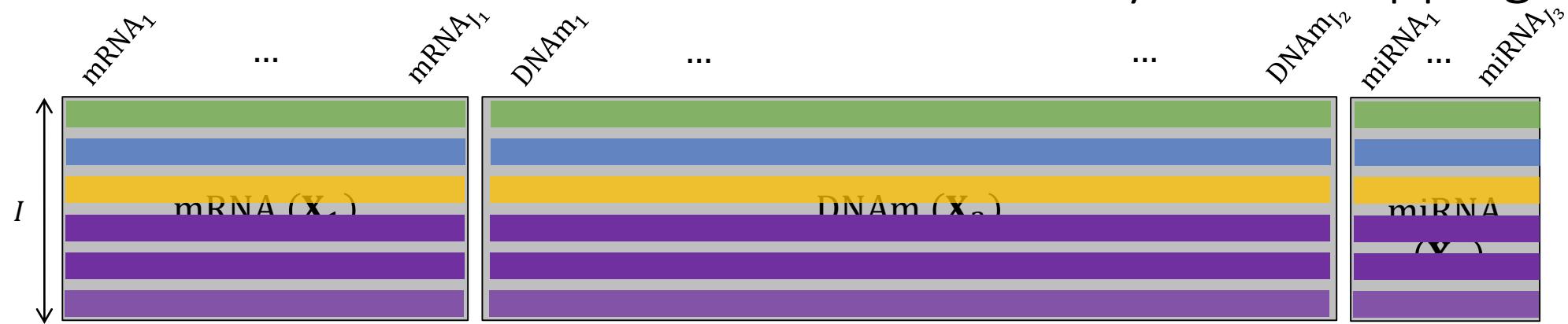
Evaluate the robustness of the model by bootstrapping.



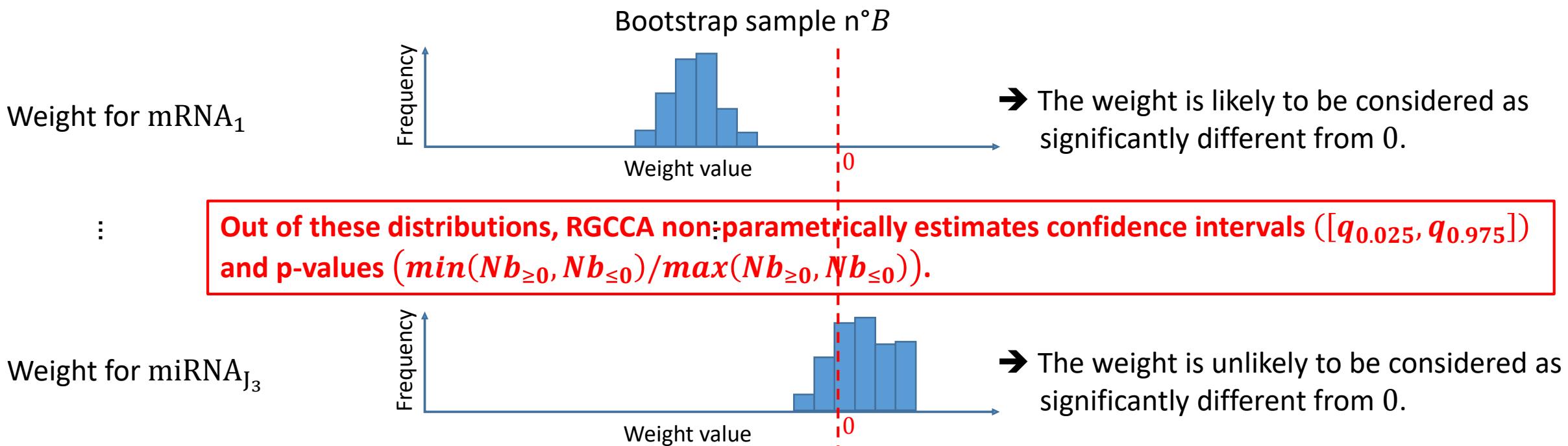
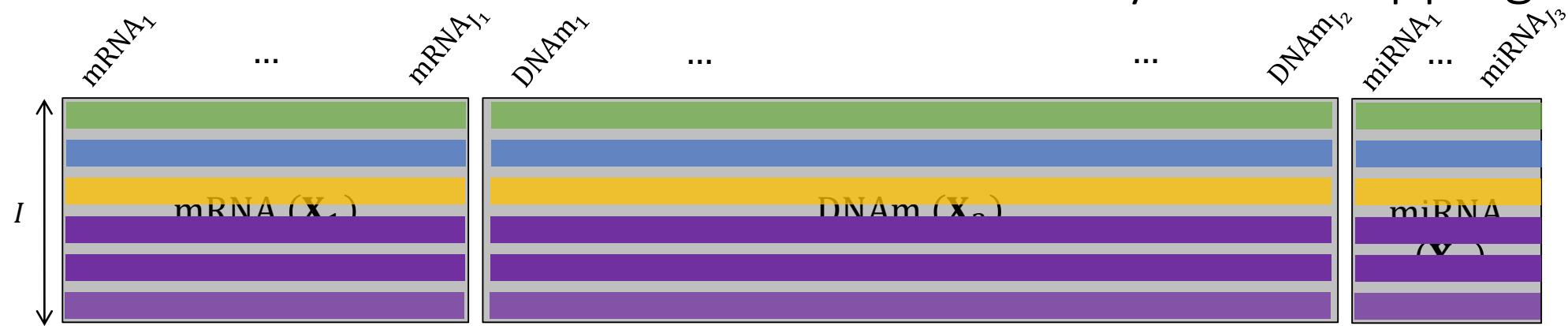
Evaluate the robustness of the model by bootstrapping.



Evaluate the robustness of the model by bootstrapping.



Evaluate the robustness of the model by bootstrapping.

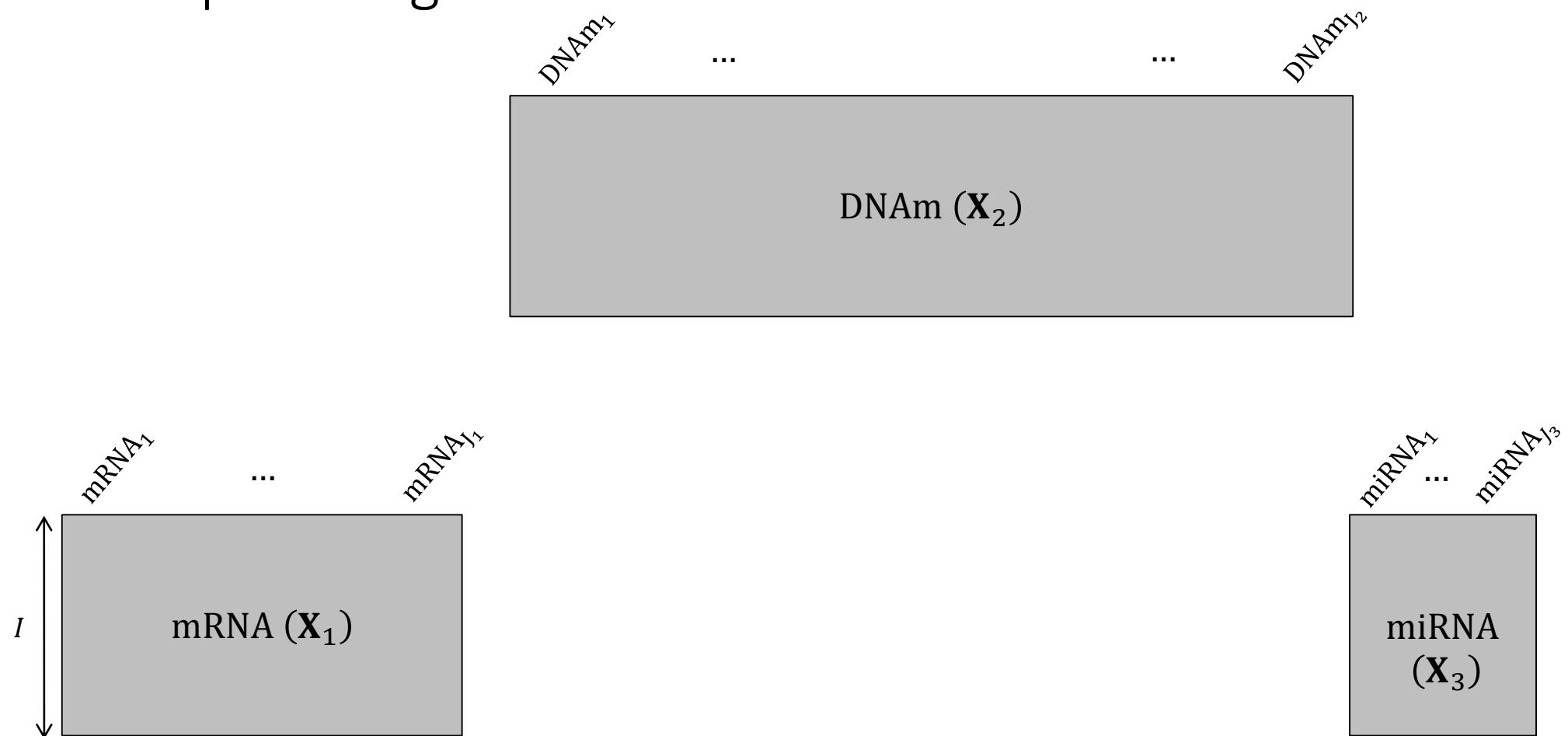




5. Supervised analysis with L -blocks

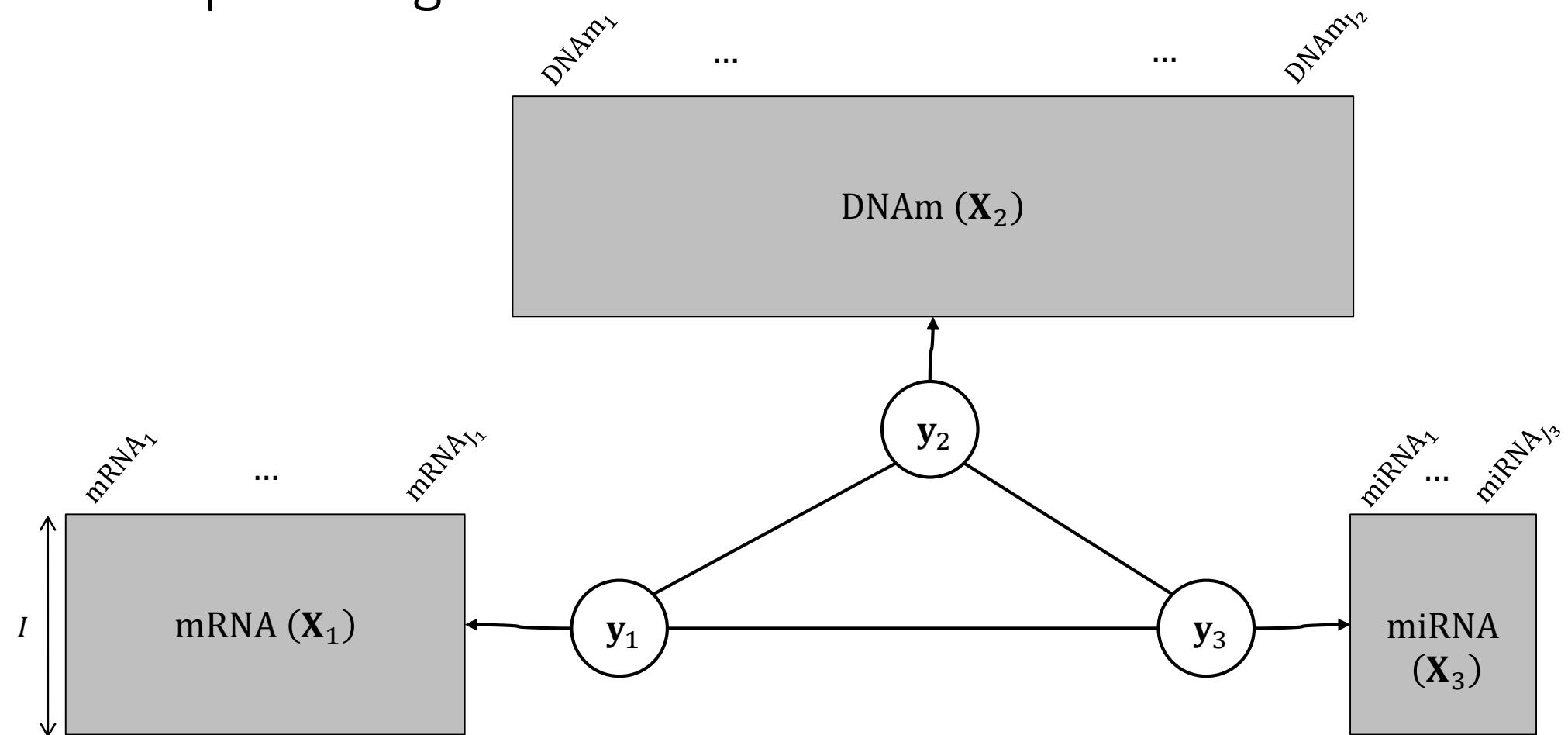


Supervising with RGCCA



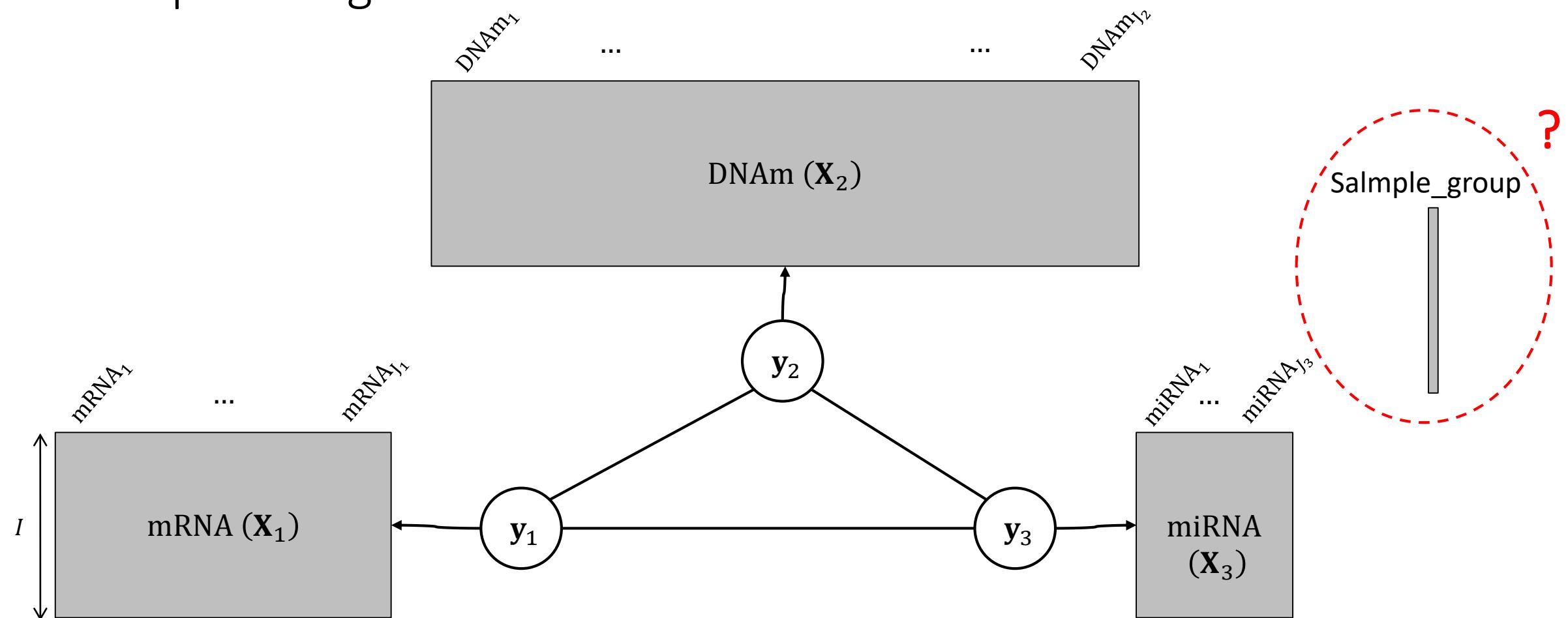


Supervising with RGCCA

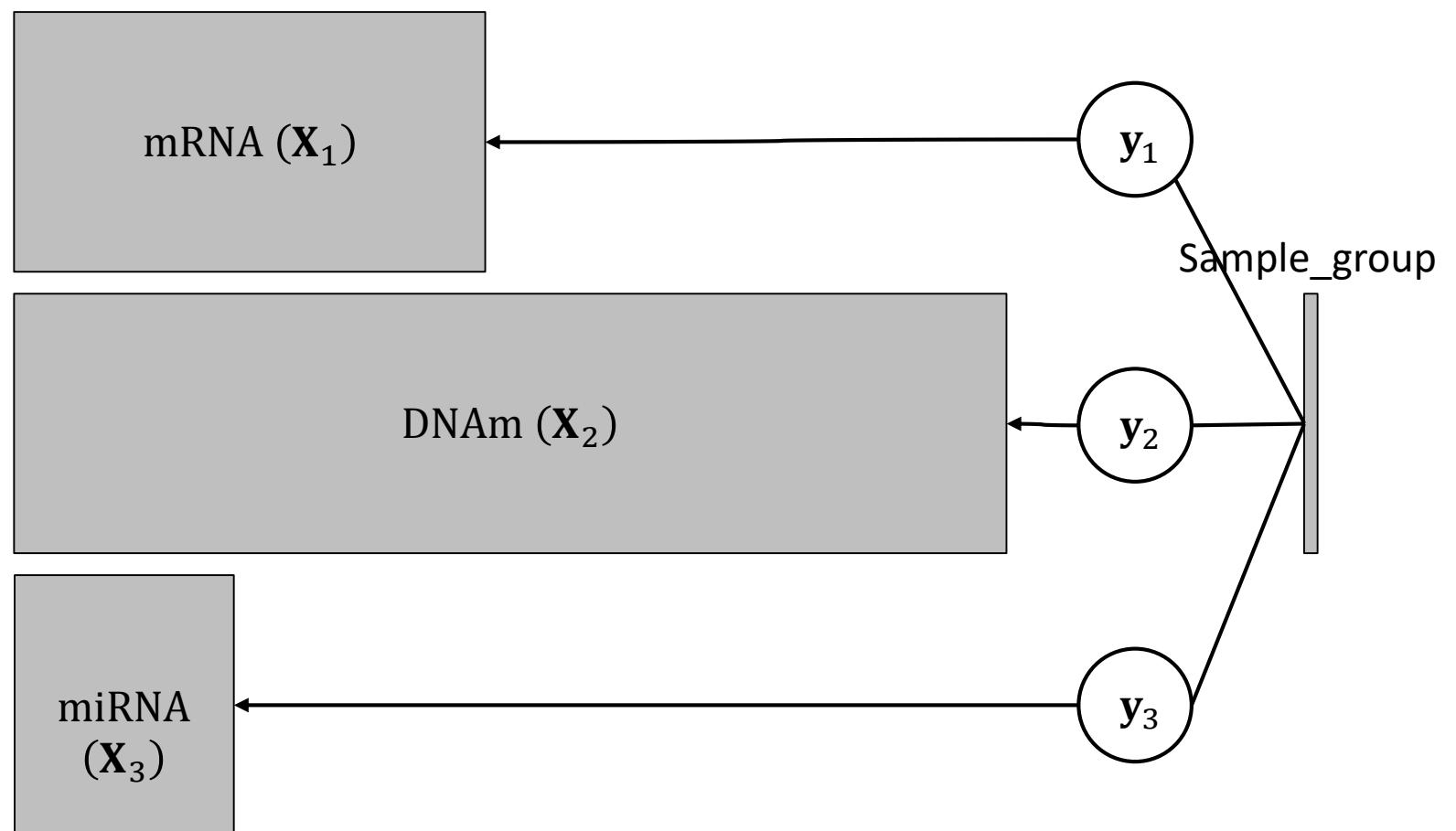




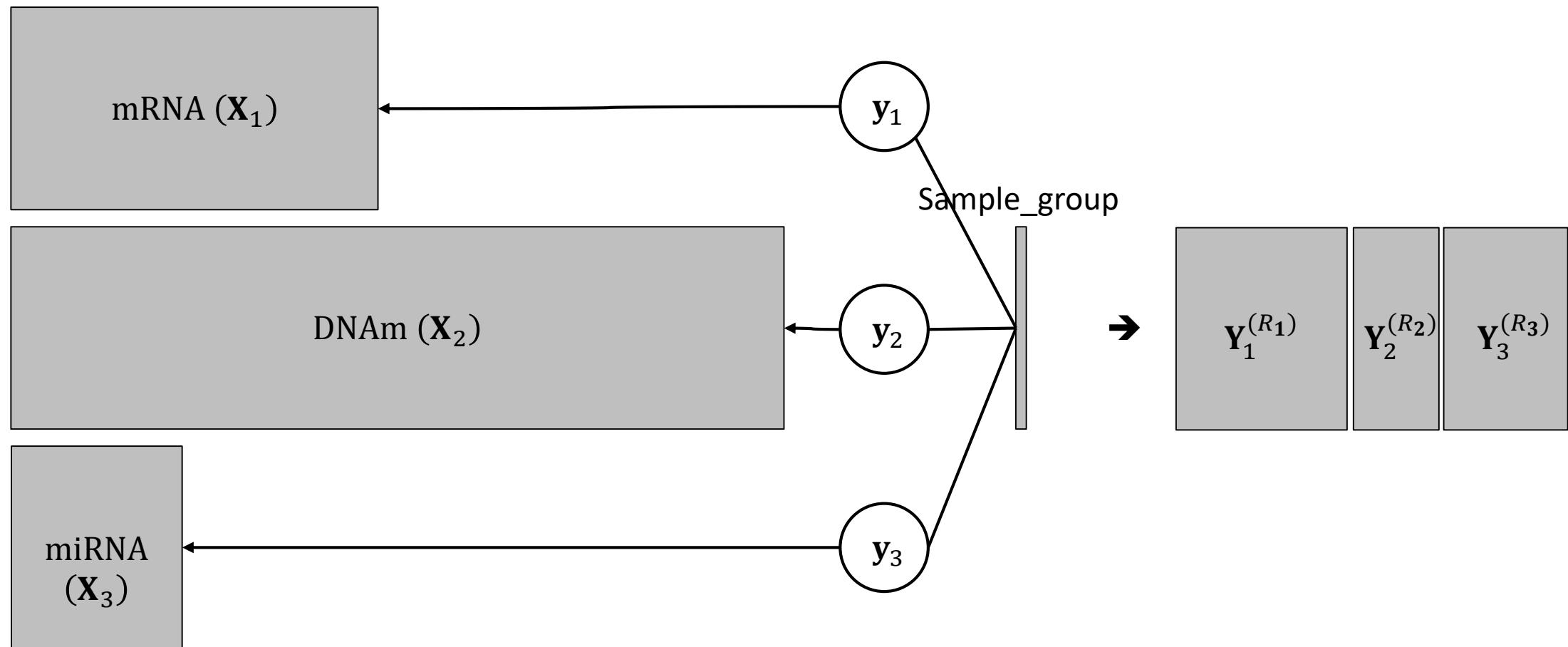
Supervising with RGCCA



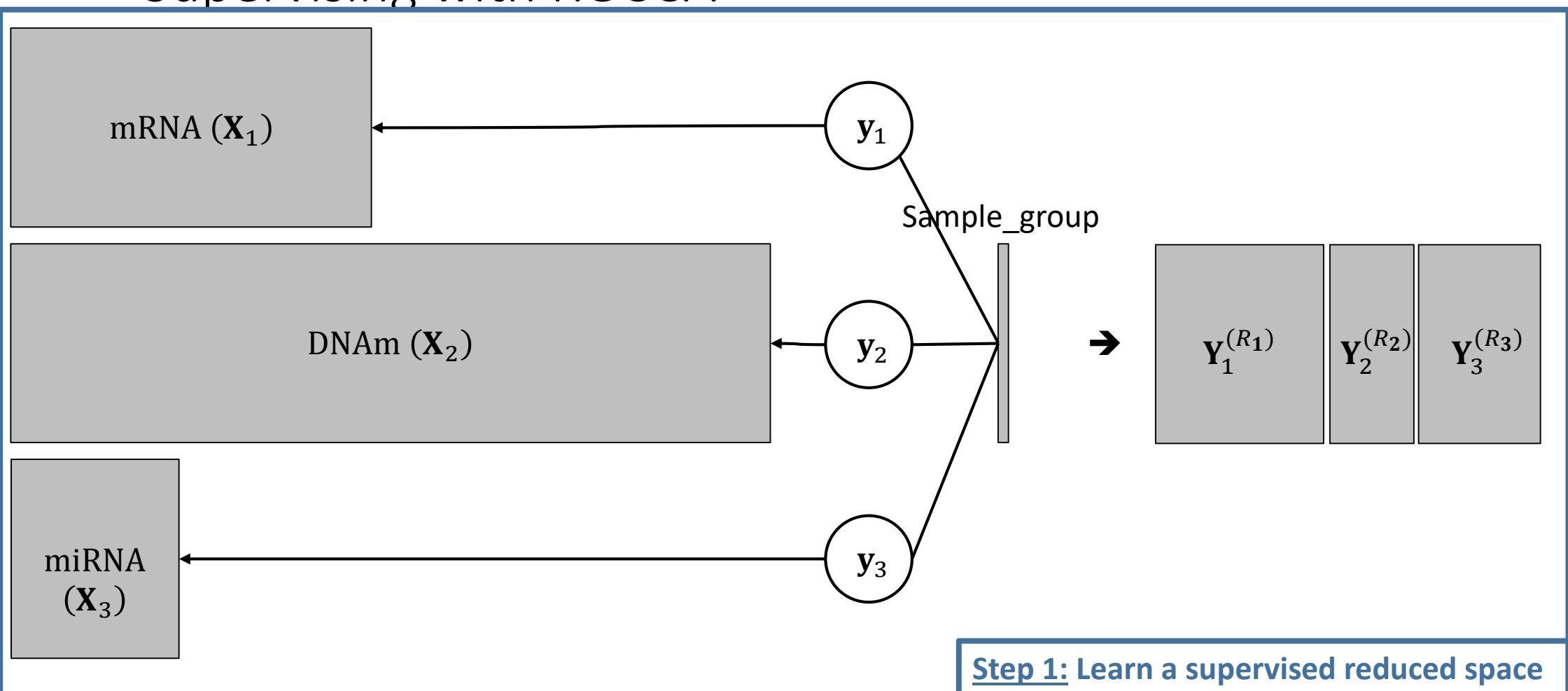
Supervising with RGCCA



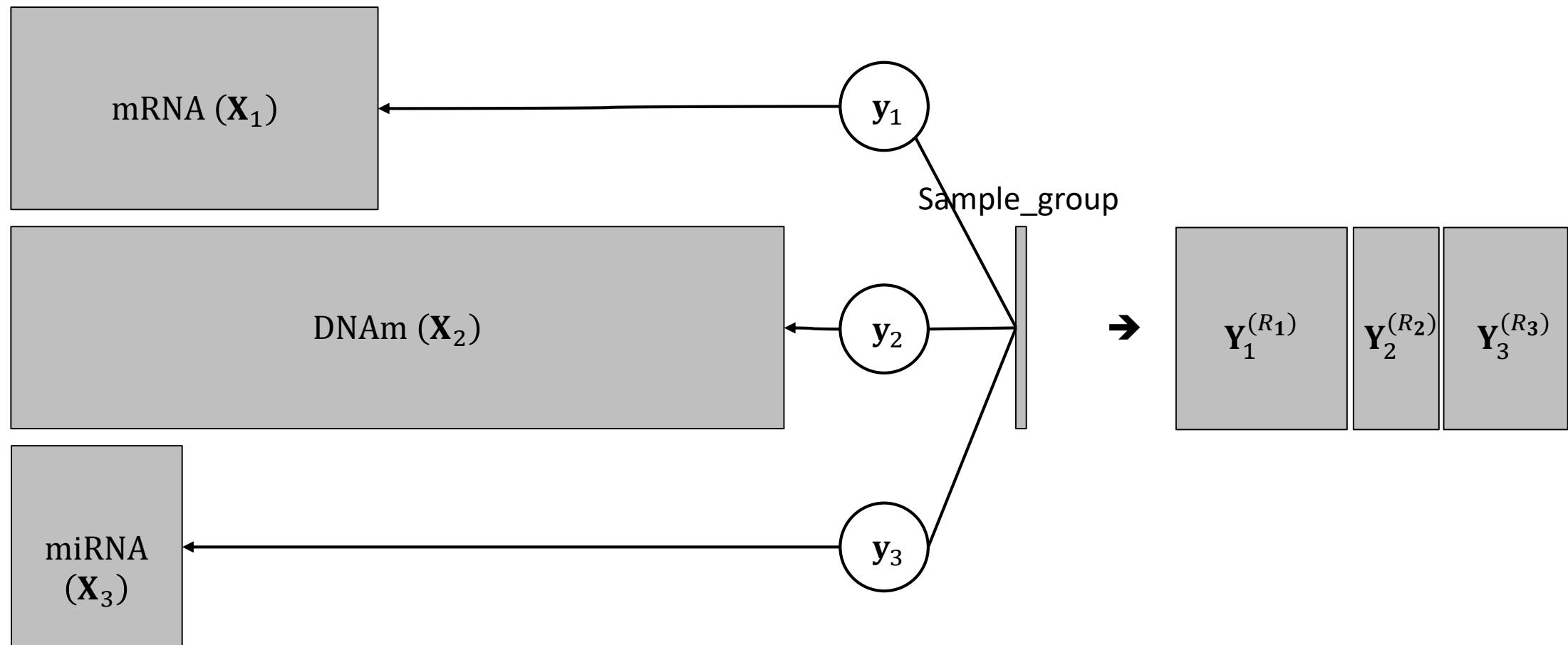
Supervising with RGCCA



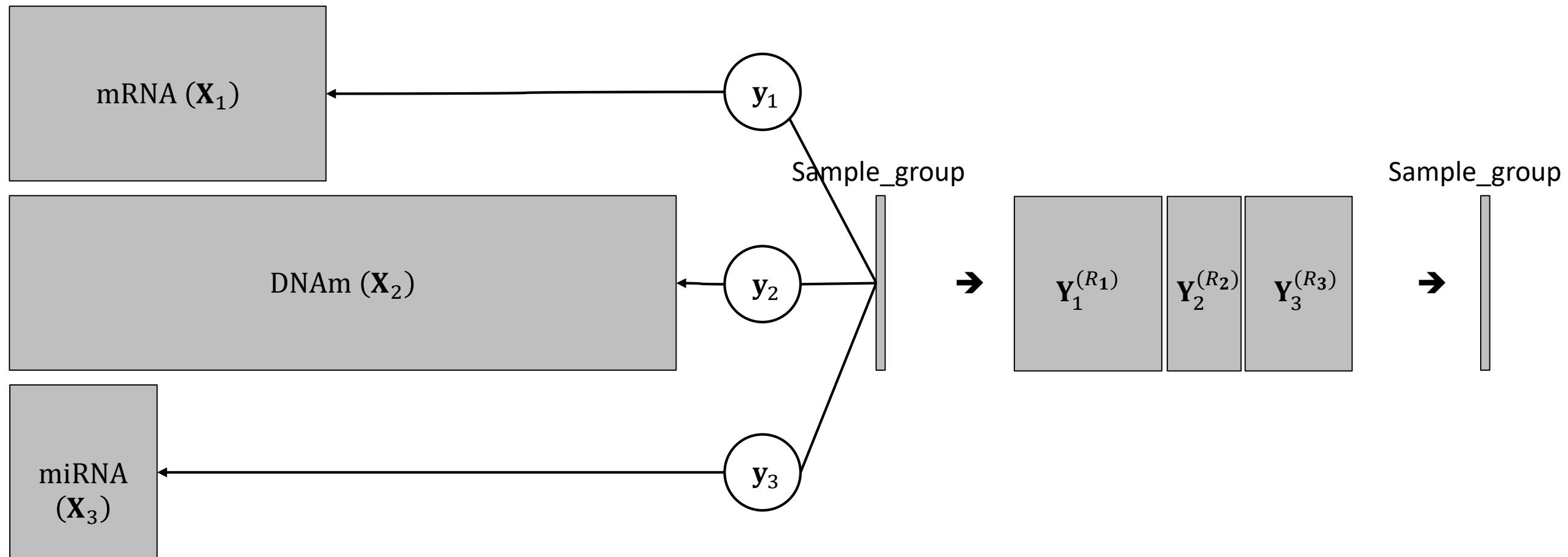
Supervising with RGCCA



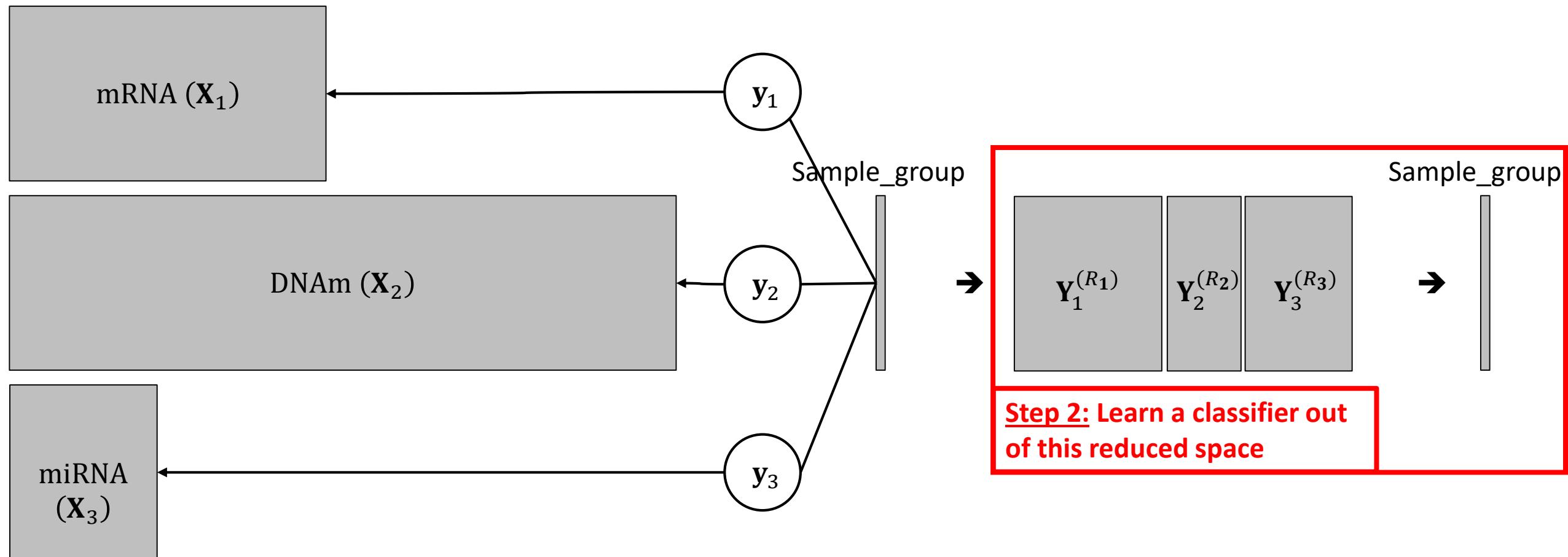
Supervising with RGCCA



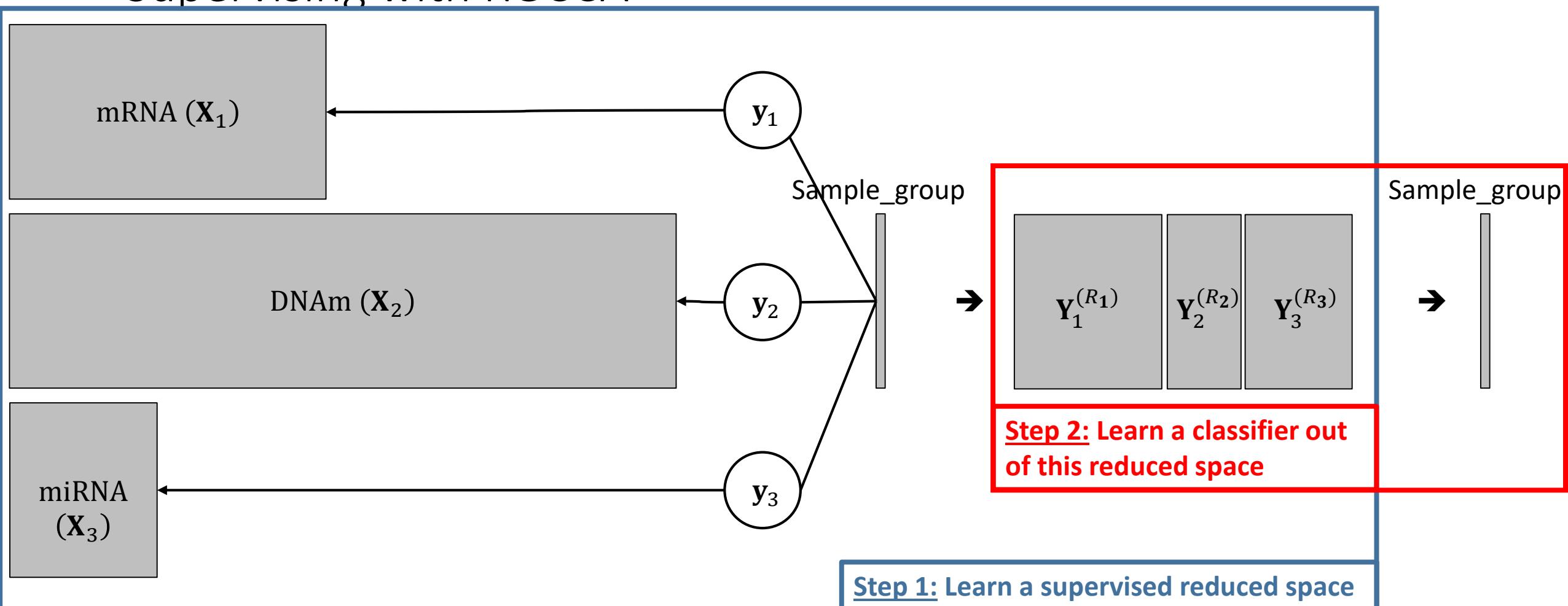
Supervising with RGCCA



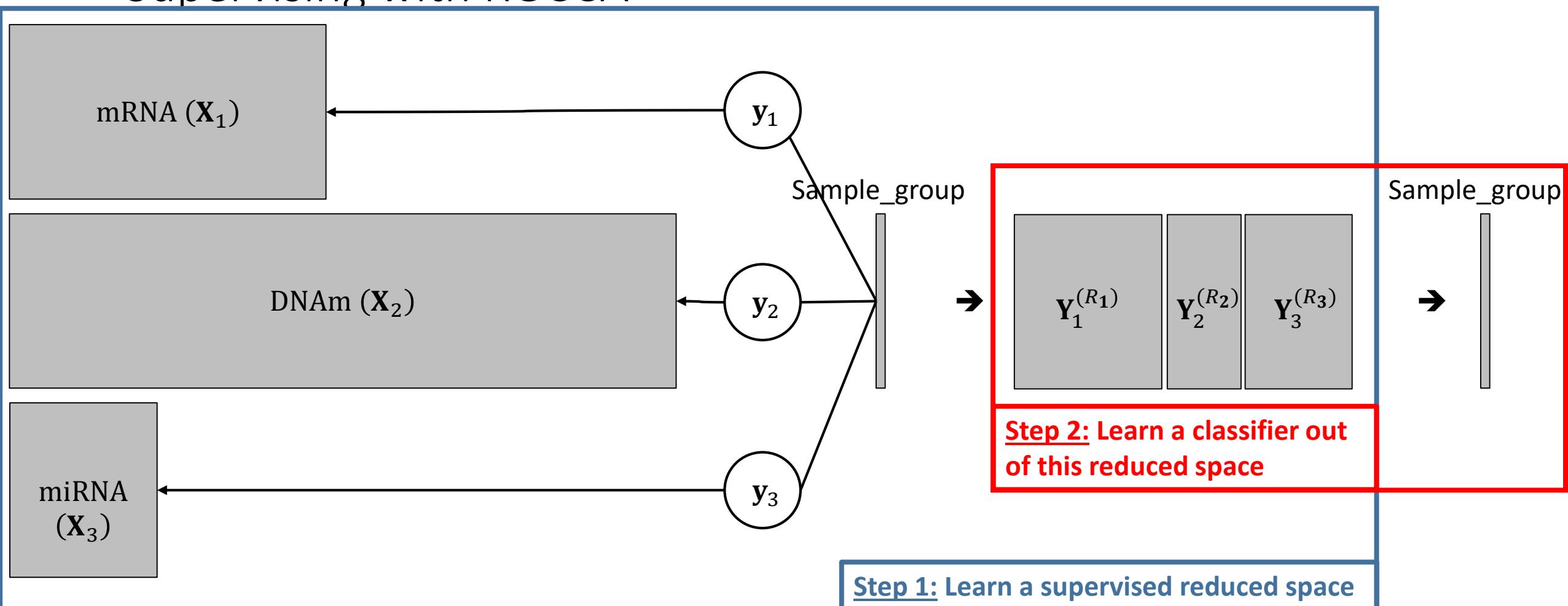
Supervising with RGCCA



Supervising with RGCCA



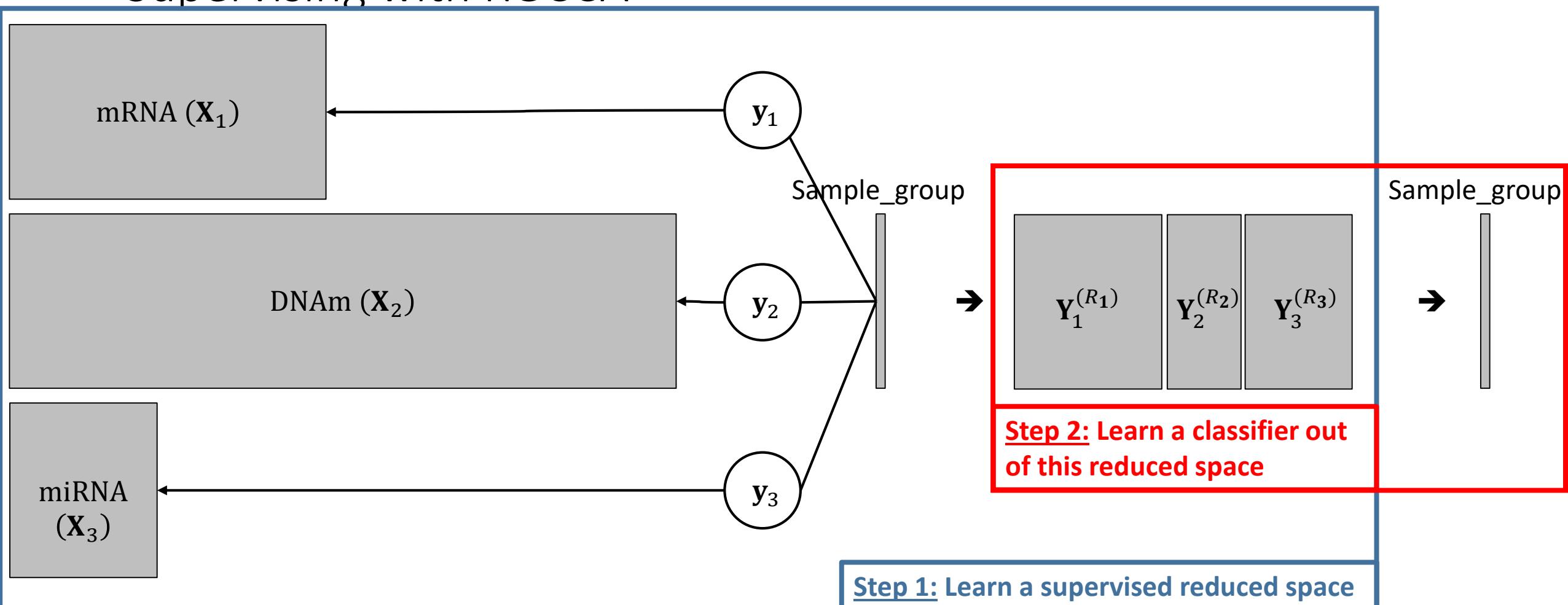
Supervising with RGCCA



→ 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 ?



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 ?

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

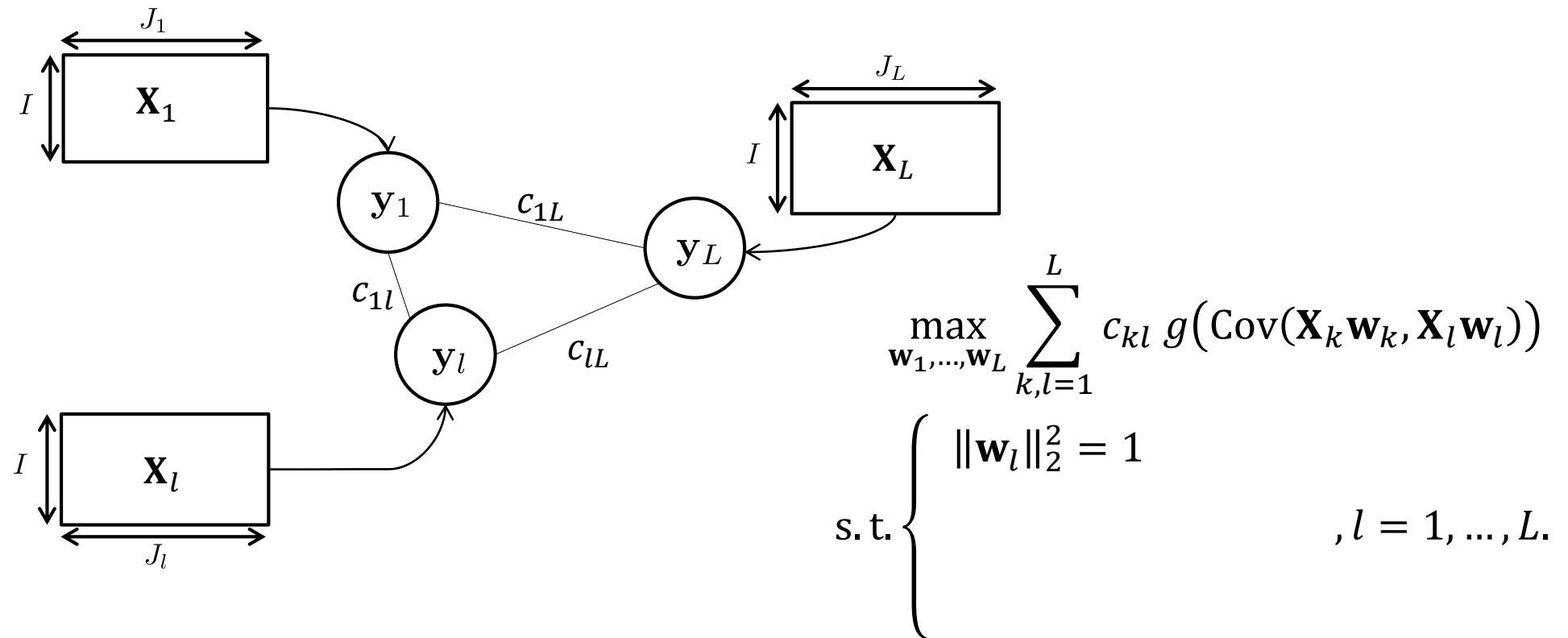
→ How many true positive labels are retrieved ?

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

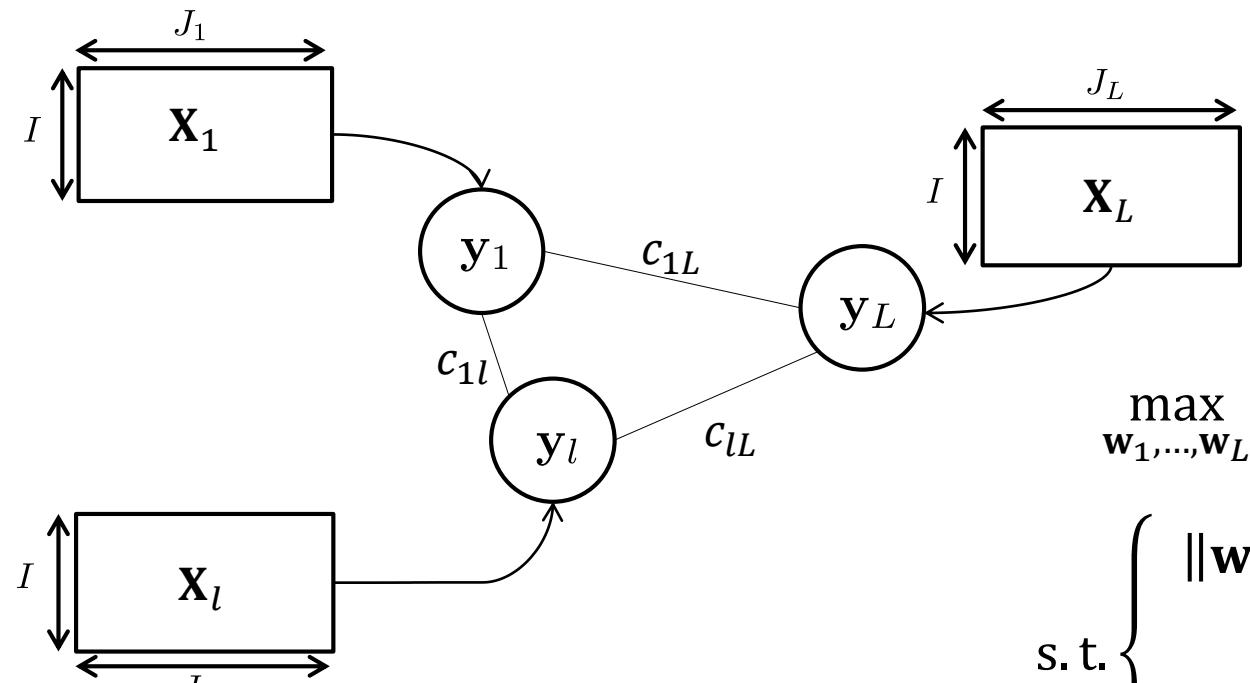


6. Going further with RGCCA

Sparse Generalized Canonical Correlation Analysis (SGCCA)



Sparse Generalized Canonical Correlation Analysis (SGCCA)

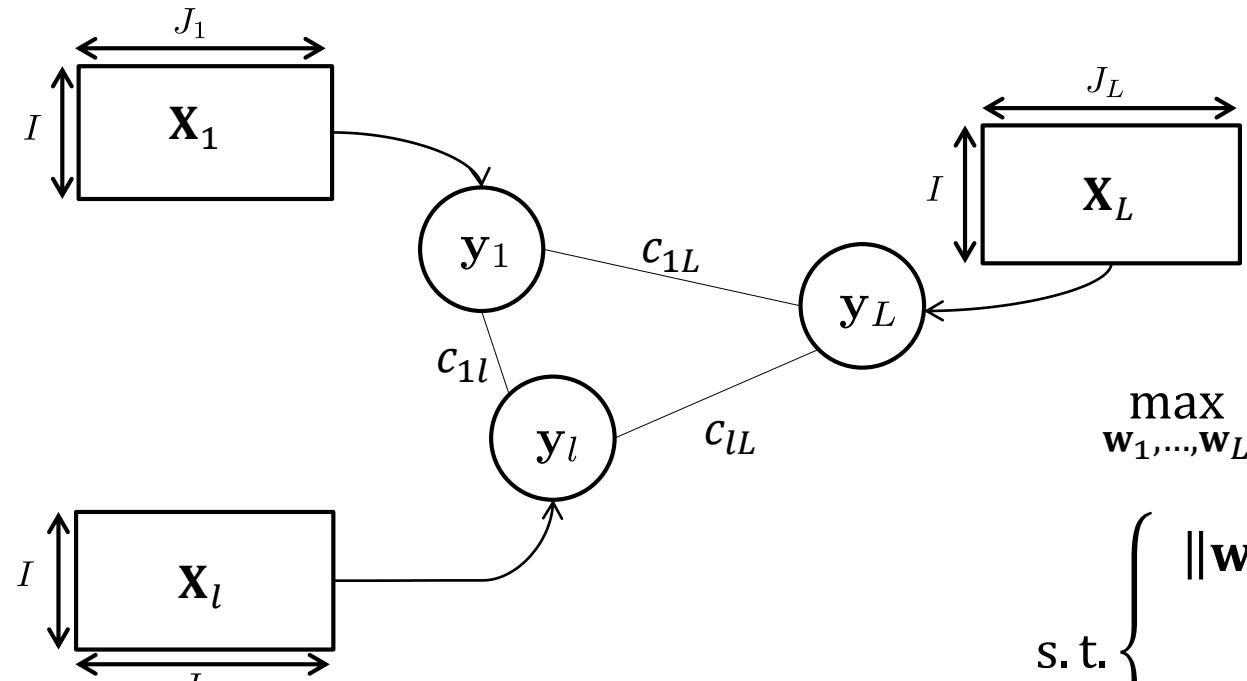


$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l))$$

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



Sparse Generalized Canonical Correlation Analysis (SGCCA)

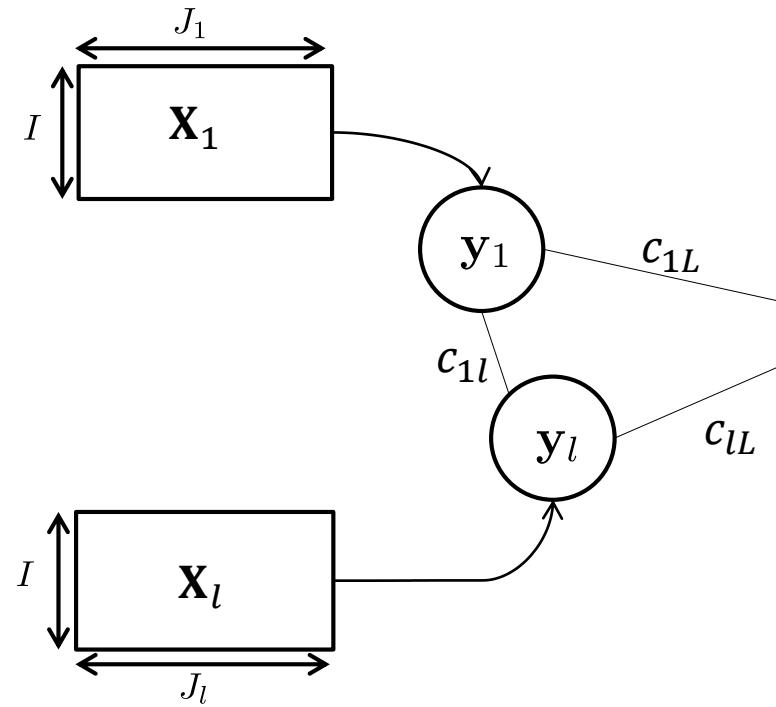


$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l))$$

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

→ The LASSO regularization allows to perform variable selection.

Sparse Generalized Canonical Correlation Analysis (SGCCA)



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

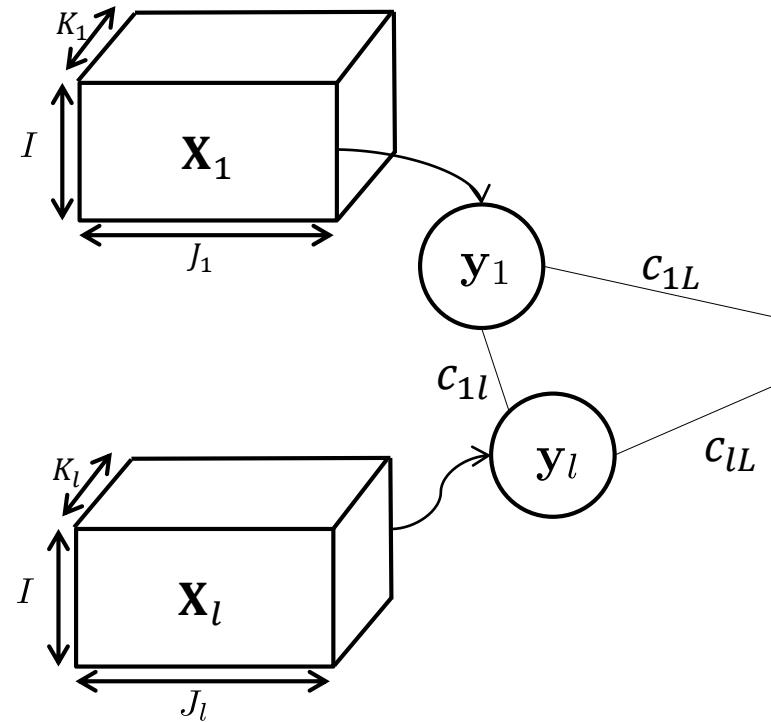
s.t.
$$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \|\mathbf{w}_l\|_1 = \sum_{j=1}^{J_l} |w_{lj}| \leq s_l \end{cases}, l = 1, \dots, L.$$

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

→ The LASSO regularization allows to perform variable selection.



Multiway Generalized Canonical Correlation Analysis (MGCCA)

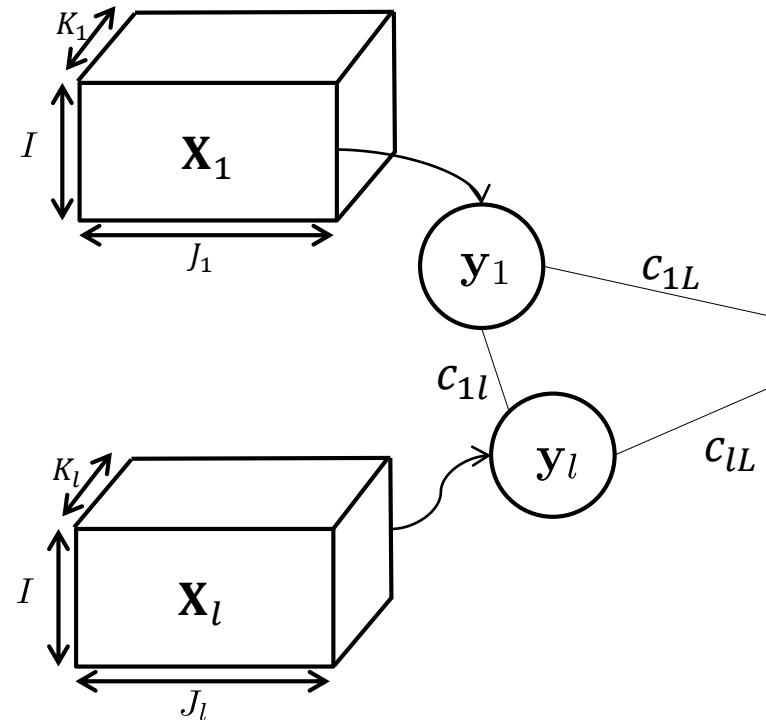


$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l))$$

$$\text{s. t. } \left\{ \begin{array}{l} \|\mathbf{w}_l\|_2^2 = 1 \\ , l = 1, \dots, L. \end{array} \right.$$



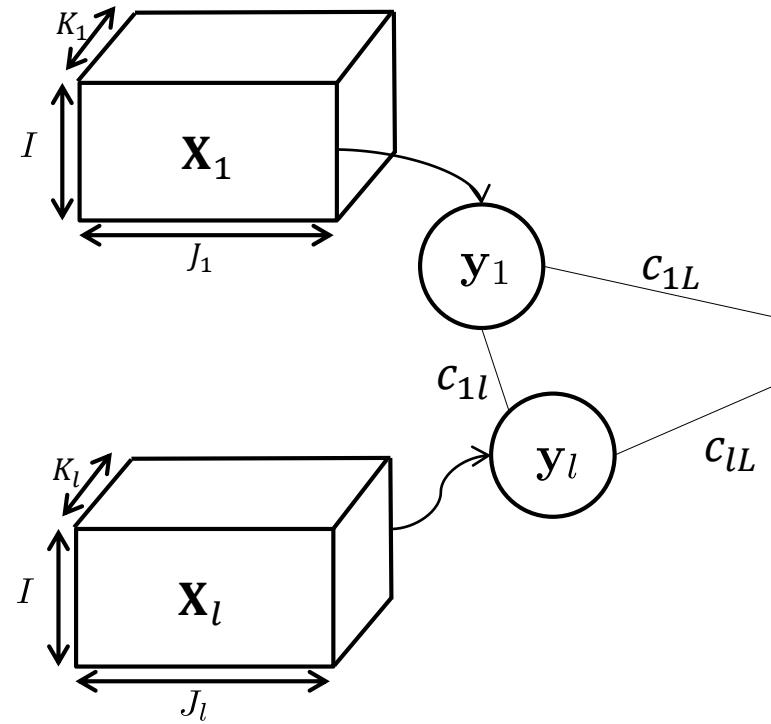
Multiway Generalized Canonical Correlation Analysis (MGCCA)



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

s.t.
$$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \dots, L.$$

Multiway Generalized Canonical Correlation Analysis (MGCCA)

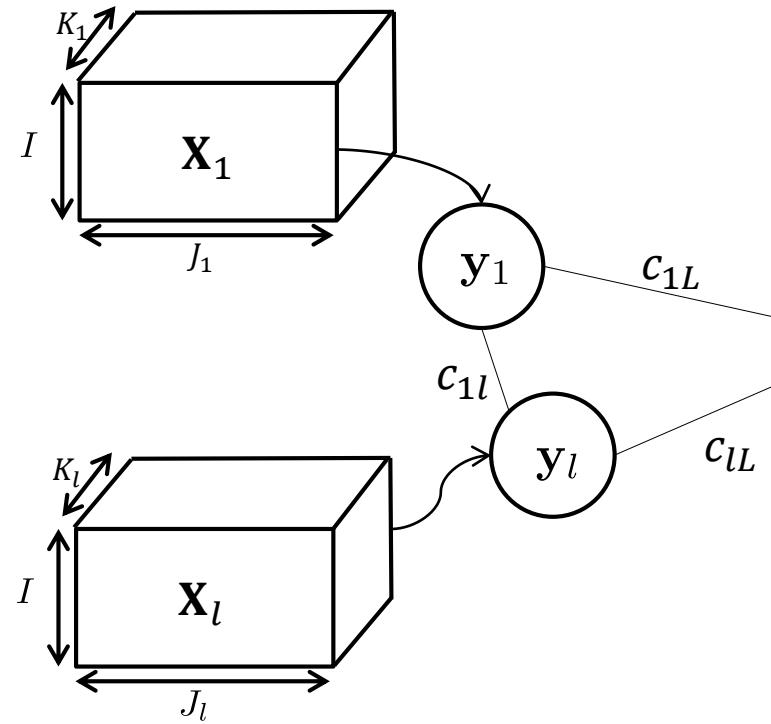


$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l))$$

s.t.
$$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \dots, L.$$



Multiway Generalized Canonical Correlation Analysis (MGCCA)

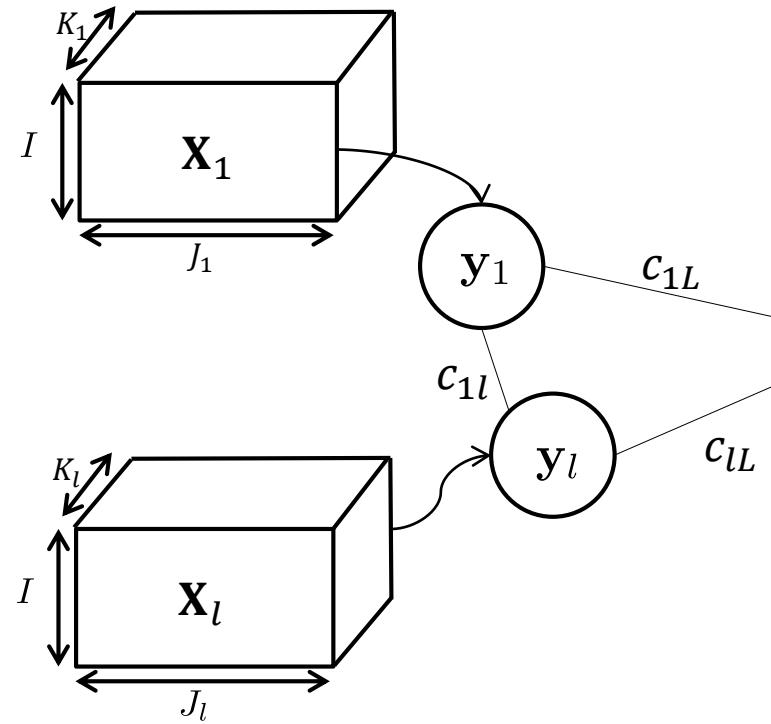


$$\max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l))$$

s.t.
$$\begin{cases} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{cases}, l = 1, \dots, L.$$

Example of such data: Electro-EncephaloGrams. Advantages:

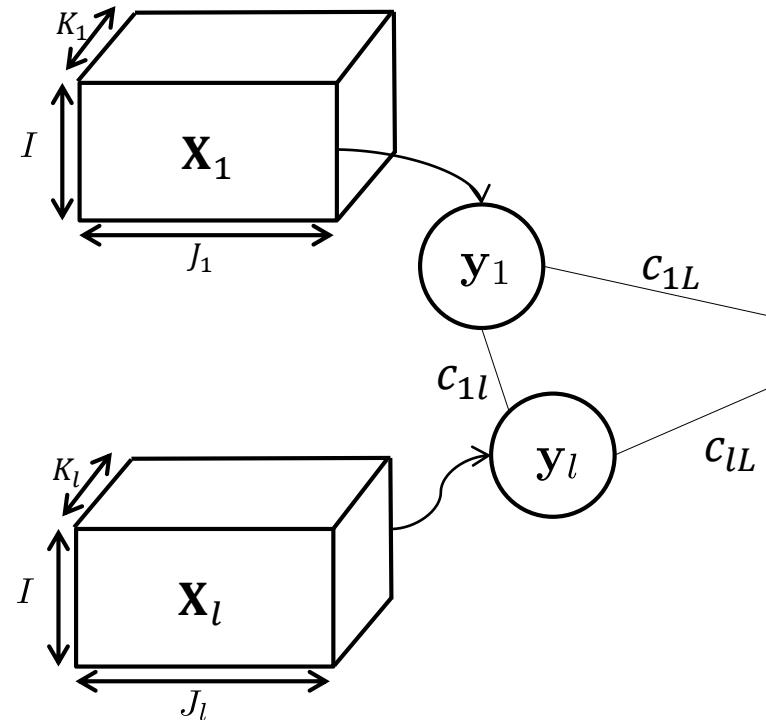
Multiway Generalized Canonical Correlation Analysis (MGCCA)



$$\begin{aligned}
 & \max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l)) \\
 \text{s.t. } & \left\{ \begin{array}{l} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{array} \right. , l = 1, \dots, L.
 \end{aligned}$$

Example of such data: Electro-EncephaloGrams. Advantages:
 1. The tensor structure is taken into account.

Multiway Generalized Canonical Correlation Analysis (MGCCA)

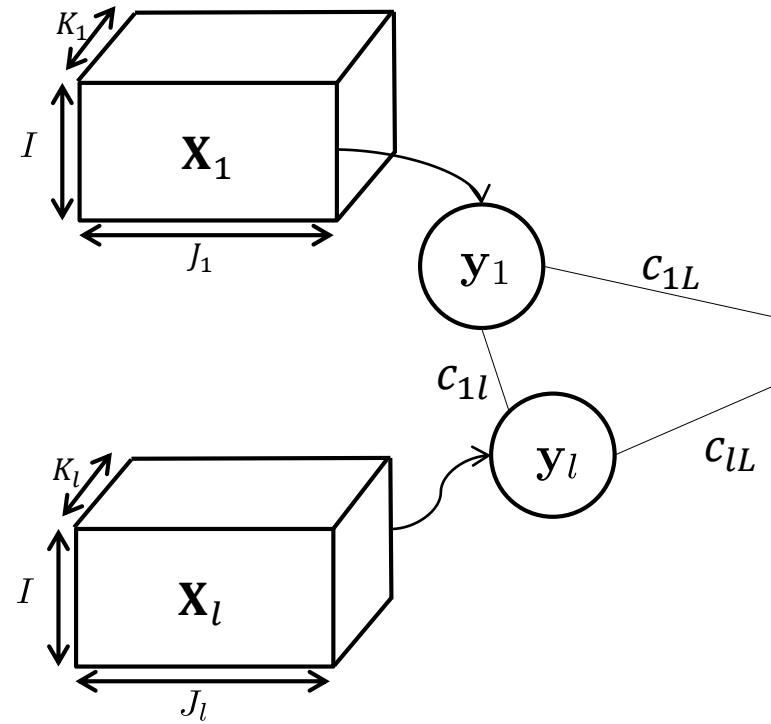


$$\begin{aligned}
 & \max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l)) \\
 \text{s.t. } & \left\{ \begin{array}{l} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{array} \right. , l = 1, \dots, L.
 \end{aligned}$$

→ Example of such data: Electro-EncephaloGrams. Advantages:

1. The tensor structure is taken into account.
2. We can separately interpret the influence of each dimension.

Multiway Generalized Canonical Correlation Analysis (MGCCA)

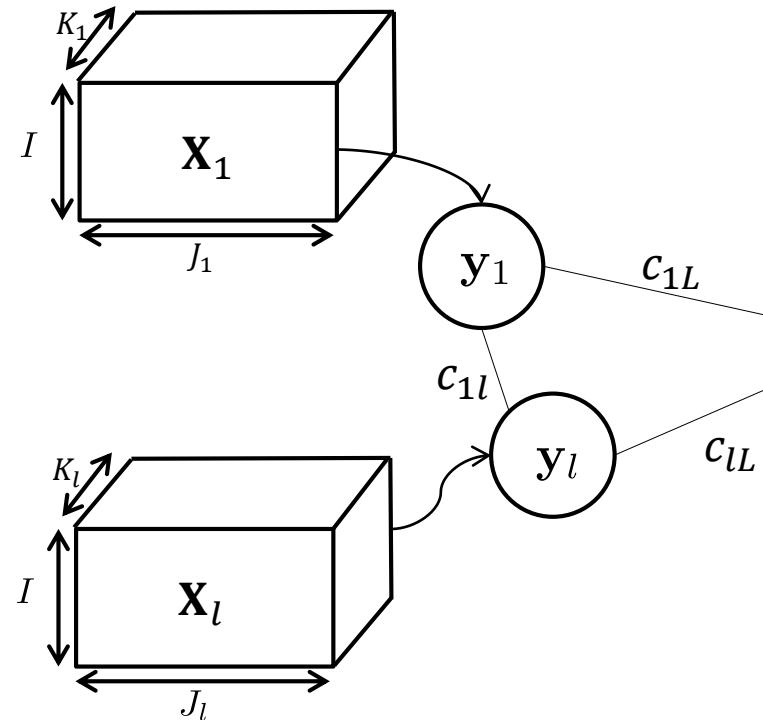


$$\begin{aligned}
 & \max_{\mathbf{w}_1, \dots, \mathbf{w}_L} \sum_{k,l=1}^L c_{kl} g(\text{Cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l)) \\
 \text{s.t. } & \left\{ \begin{array}{l} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{array} \right. , l = 1, \dots, L.
 \end{aligned}$$

→ Example of such data: Electro-EncephaloGrams. Advantages:

1. The tensor structure is taken into account.
2. We can separately interpret the influence of each dimension.
3. The number of parameters to estimate can be tremendously reduced (from $J_l \times K_l$ to $J_l + K_l$).

Multiway Generalized Canonical Correlation Analysis (MGCCA)



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

s.t. $\left\{ \begin{array}{l} \|\mathbf{w}_l\|_2^2 = 1 \\ \mathbf{w}_l = \mathbf{w}_l^K \otimes \mathbf{w}_l^J \end{array} \right. , l = 1, \dots, L.$

Example of such data: Electro-EncephaloGrams. Advantages:

1. The tensor structure is taken into account.
2. We can separately interpret the influence of each dimension.
3. The number of parameters to estimate can be tremendously reduced (from $J_l \times K_l$ to $J_l + K_l$).

New extension with Tensor GCCA



Other extensions of RGCCA



Other extensions of RGCCA

Kernel GCCA: in order to take estimate non-linear links between blocks.



Other extensions of RGCCA

Kernel GCCA: in order to take estimate non-linear links between blocks.

Functional GCCA: in order to handle longitudinal blocks.



Other extensions of RGCCA

Kernel GCCA: in order to take estimate non-linear links between blocks.

Functional GCCA: in order to handle longitudinal blocks.

Multi-group: find relationships between variables within each group that are common to all groups.



Other extensions of RGCCA

Kernel GCCA: in order to take estimate non-linear links between blocks.

Functional GCCA: in order to handle longitudinal blocks.

Multi-group: find relationships between variables within each group that are common to all groups.

Graph-net SGCCA: add a regularization term based on a gene network.



Other extensions of RGCCA

Kernel GCCA: in order to take estimate non-linear links between blocks.

Functional GCCA: in order to handle longitudinal blocks.

Multi-group: find relationships between variables within each group that are common to all groups.

Graph-net SGCCA: add a regularization term based on a gene network.

Group SGCCA: perform group selection.

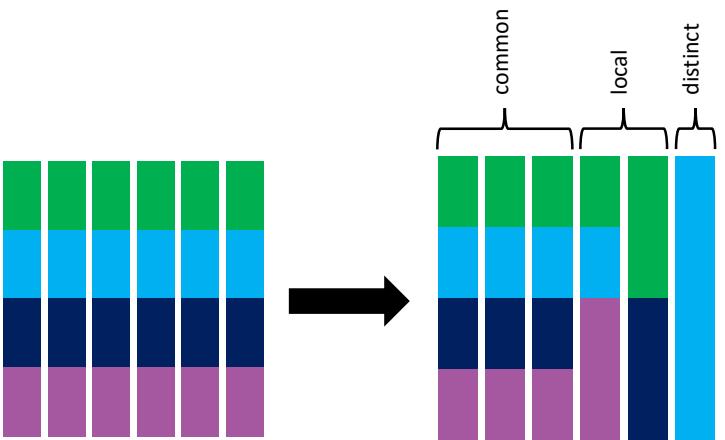


Research Axes



Research Axes

Axe 1: Use common and specific information

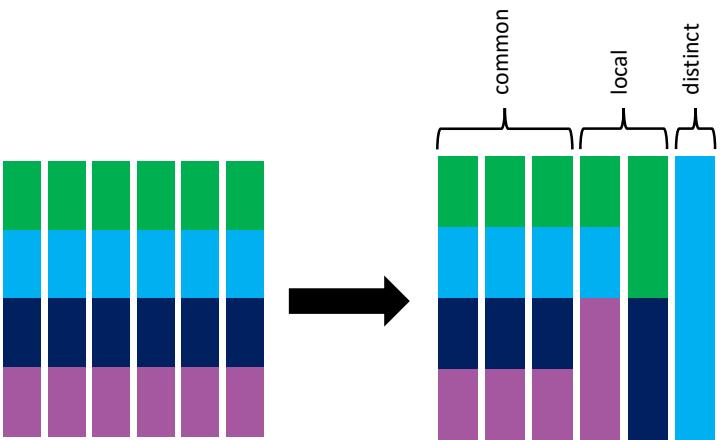




Research Axes

Axe 1: Use common and specific information

Vary the combination of omics data from which components are built.



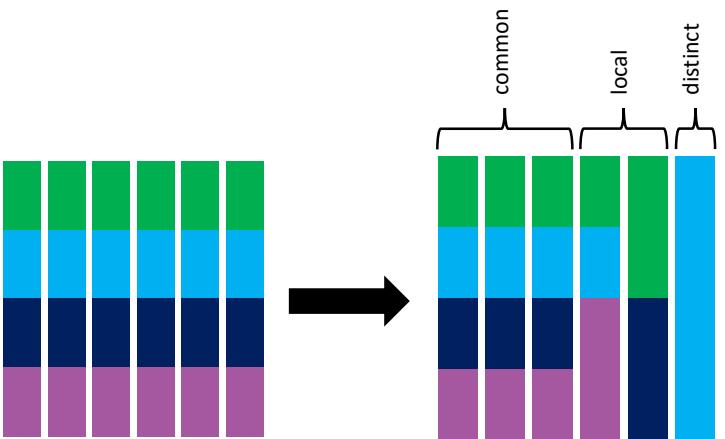


Research Axes

Axe 1: Use common and specific information

Vary the combination of omics data from which components are built.

Improve the interpretation of components.



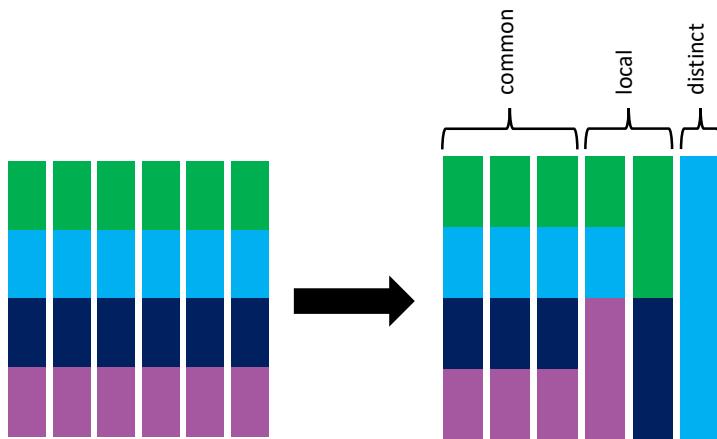


Research Axes

Axe 1: Use common and specific information

Vary the combination of omics data from which components are built.

Improve the interpretation of components.



Axe 2: Include the appartenence of each variable to a biological pathway.

Divide each omic matrix by biological pathways.

Allow to identify most important pathways (With L1 norm).



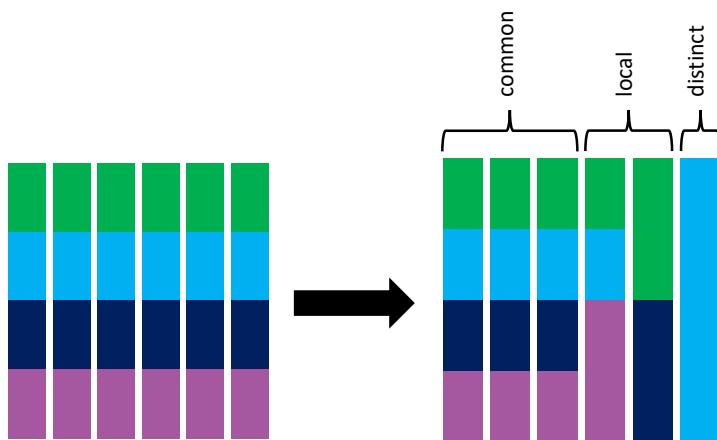


Research Axes

Axe 1: Use common and specific information

Vary the combination of omics data from which components are built.

Improve the interpretation of components.



Axe 2: Include the appartenence of each variable to a biological pathway.

Divide each omic matrix by biological pathways.

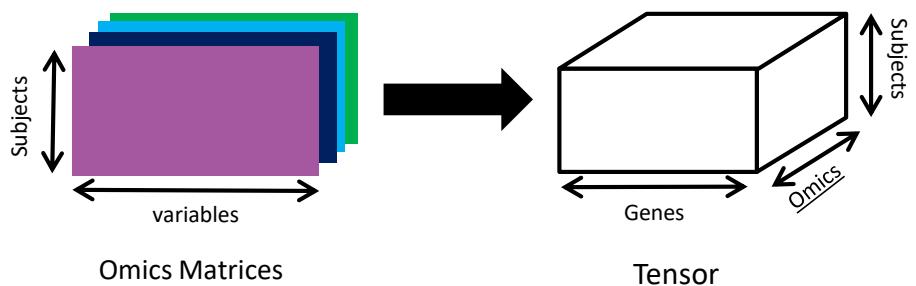
Allow to identify most important pathways (With L1 norm).



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-Scar 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 (V1):

Florence Mehl
Olivier Sand
Grégoire Rossier
Jimmy Vandel
Guillemette Marot
Marie-Galadriel Briere
Morgane Terezol
Anaïs Baudot
Maxime Delmas
Jean-Clément Gallardo
Marco Pagni

Statomqiue (V2):

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



QUESTIONS ?