



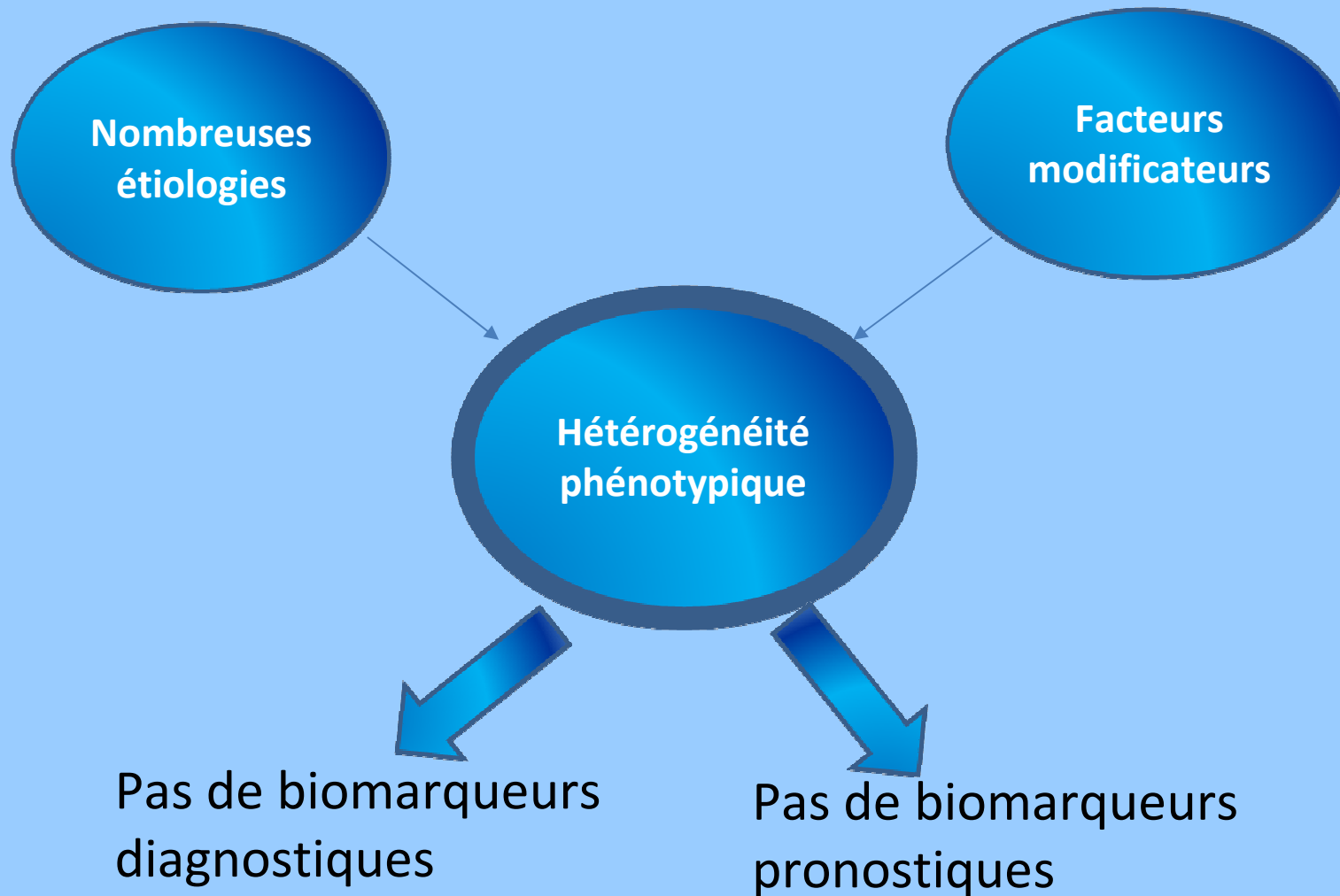
*Comparaison d'une méthode d'analyse multivariée OPLS-DA et d'une méthode de fouille de données pour l'identification de biomarqueurs de Sclérose Latérale Amyotrophique par approche métabolomique*

**H. Blasco**, J. Błaszczyn'ski, **J.C. Billaut**, L. Nadal-Desbarats, P.F. Pradat, D. Devos, C. Moreau, C.R. Andres, P. Emond, P. Corcia, R. Słowin'ski

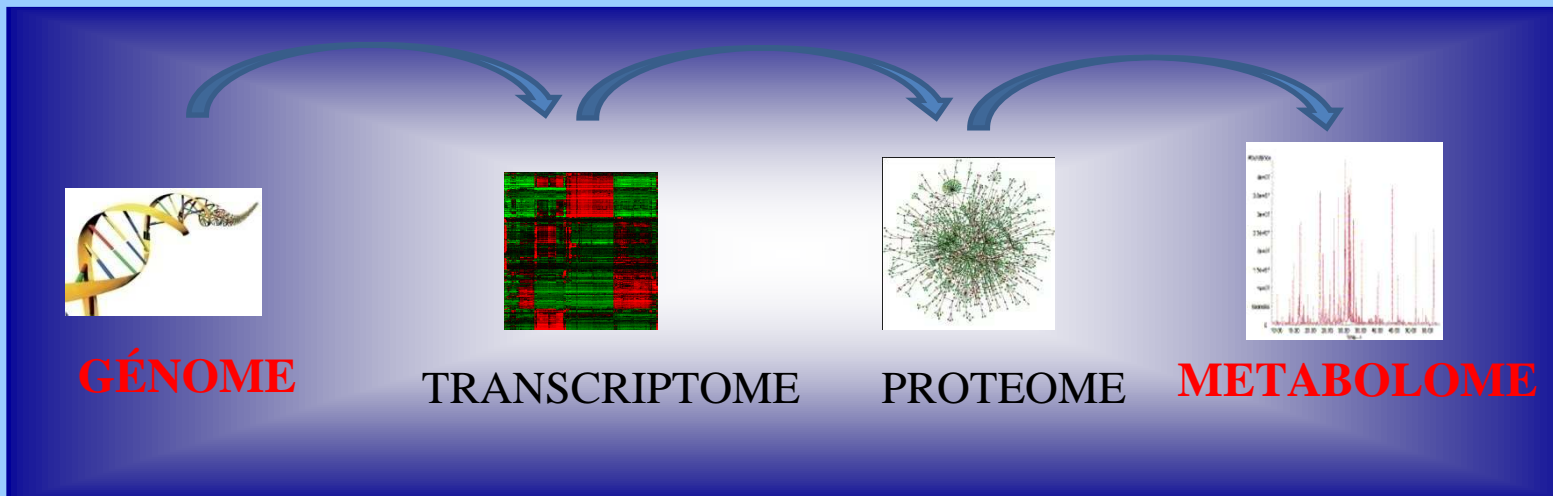
*Journal of Biomedical Informatics, 2015*

**2e journée du projet CASCIMODOT  
8 juillet 2015**

# La SLA : une maladie hétérogène

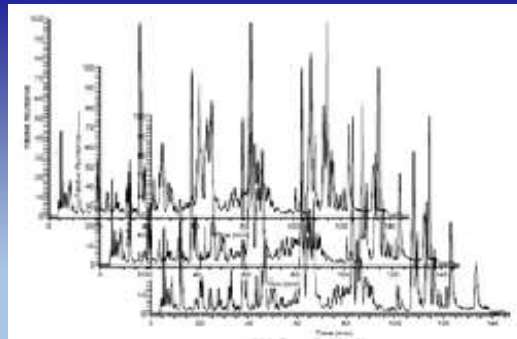


# La promesse de la métabolomique



# L'analyse métabolomique

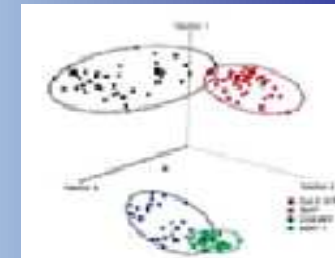
Etape pré-analytique



(SRM et SM)

Empreintes métaboliques

Traitement des données  
-Suppression du bruits de fond  
-Alignement des pics



Analyse  
statistique  
multivariée



Pertinence  
biologique



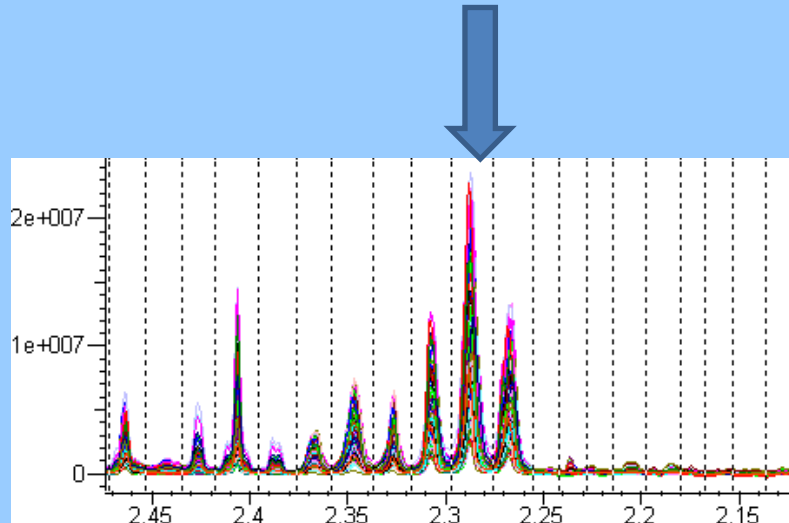
Identification  
(SRM, SM)



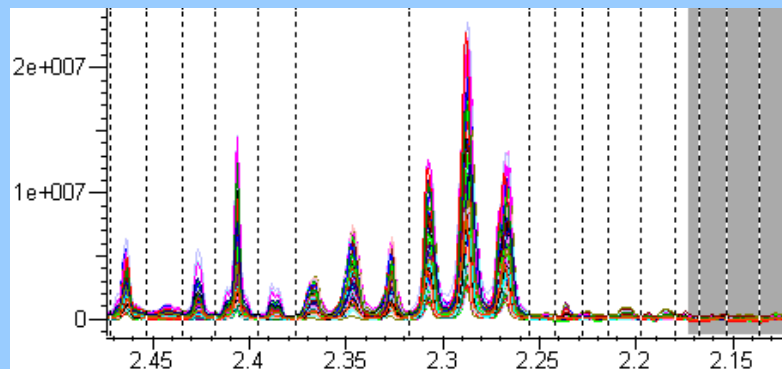
Métabolites  
d'intérêt

# Traitement des données : découpage du spectre

Données spectrales → matrice de données



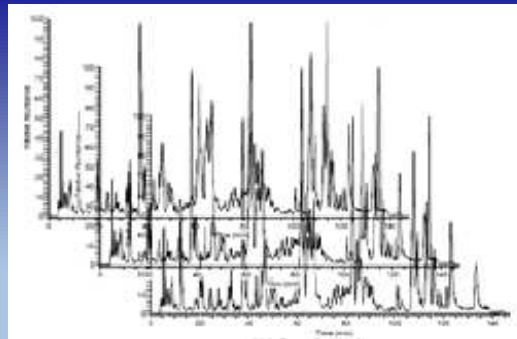
500 variables



100 variables

# L'analyse métabolomique

Etape pré-analytique



(SRM et SM)

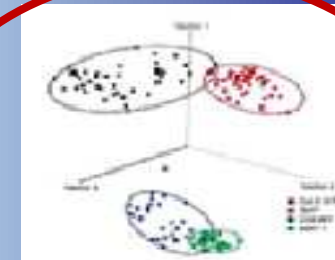
Empreintes métaboliques



Traitement des données  
-Suppression du bruit de fond  
-Alignement des pics



Analyse  
statistique  
multivariée



Pertinence  
biologique



Identification  
(SRM, SM)



Métabolites  
d'intérêt

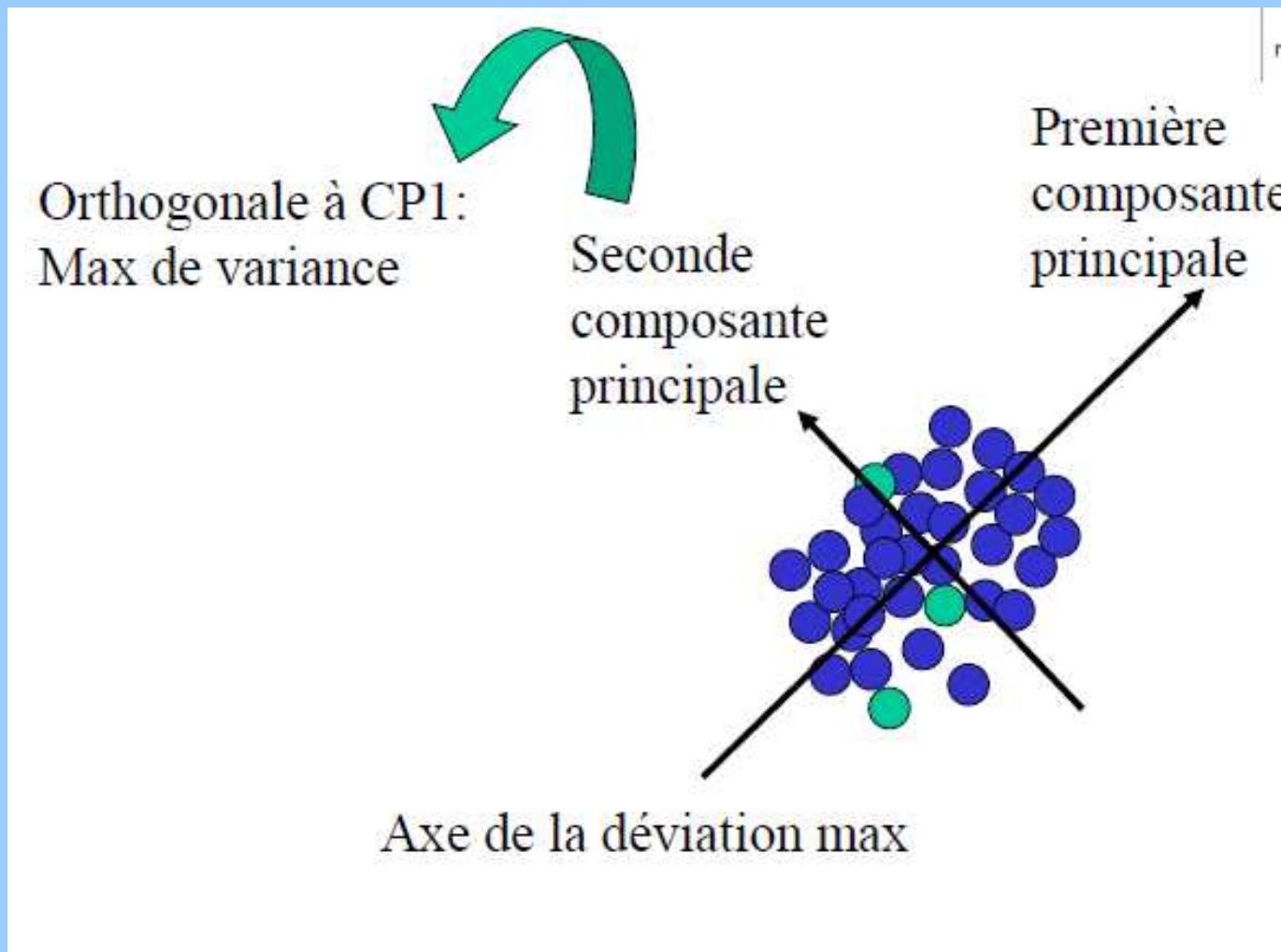
# Traitement des données

## Analyse non supervisée

- Pas besoin de connaissance *a priori* sur les échantillons
- Utilisée en première approche, notamment pour repérer les échantillons aberrants qui seront exclus et pour voir les tendances
- Mesure de la valeur intrinsèque de variabilité au sein d'un jeu de données
- Nombreuses analyses possibles : Hierarchical cluster analysis, analyse en composante principales

# Traitement des données

## Analyse non supervisée en composante principale (ACP)





# Analyse supervisée

## Partial Least Square-Discriminant Analysis (OPLS-DA)

- Appartenance des échantillons à une classe connue
- Construction de modèles robustes permettant de repérer les variables spectrales discriminantes entre les groupes
- Modèle optimisé pour être prédictif ( $R^2$ = variabilité,  $Q^2$ = prédiction)
- Cross validation (1/7 → incertitude Jack-Knife)
- 200 permutations (méthode Monte Carlo)

# Problématique de l'analyse des données métabolomiques

- Etape complexe nécessitant une expertise spécifique
- Consensus sur la méthodologie d'analyse statistique mais tendance à la remise en question de ces méthodes trop optimistes
- Nécessiter de validation robuste pour une utilisation en pratique de routine
  - cohorte indépendante
  - méthode mathématique différente

# Rough Set Approach to Knowledge Discovery about Preferences

Roman Słowiński

Poznań University of Technology  
Polish Academy of Sciences



# Quelques références

- Greco, S., Matarazzo, B., Słowiński, R.: *Rough sets theory for multi-criteria decision analysis*. *European Journal of Operational Research*, 129, 1 (2001) 1–47
- Greco, S., Matarazzo, B., Słowiński, R.: *Multicriteria classification by dominance-based rough set approach*. In: W.Kloesgen and J.Zytkow (eds.), *Handbook of Data Mining and Knowledge Discovery*, Oxford University Press, New York, 2002
- Słowiński, R., Greco, S., Matarazzo, B.: *Rough set based decision support*. Chapter 16 [in]: E.K. Burke and G. Kendall (eds.), *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, Springer-Verlag, New York (2005) 475–527
- Stefanowski, J.: *On rough set based approach to induction of decision rules*. In Skowron, A., Polkowski, L. (eds.): *Rough Set in Knowledge Discovering*, Physica Verlag, Heidelberg (1998) 500–529
- Greco S., Matarazzo, B., Słowiński, R., Stefanowski, J.: *An Algorithm for Induction of Decision Rules Consistent with the Dominance Principle*. In W. Ziarko, Y. Yao (eds.): *Rough Sets and Current Trends in Computing*. *Lecture Notes in Artificial Intelligence 2005* (2001) 304–313. Springer-Verlag
- Greco, S., B. Matarazzo, R. Slowinski and J. Stefanowski: *Variable consistency model of dominance-based rough set approach*. In W.Ziarko, Y.Yao (eds.): *Rough Sets and Current Trends in Computing*. *Lecture Notes in Artificial Intelligence 2005* (2001) 170–181. Springer-Verlag
- Dembczyński, K., Greco, S., Kotłowski, W., Słowiński, R.: *Statistical model for rough set approach to multicriteria classification*. In Kok, J.N., Koronacki, J., de Mantaras, R.L., Matwin, S., Mladenic, D., Skowron, A. (eds.): *Knowledge Discovery in Databases: PKDD 2007*, Warsaw, Poland. *Lecture Notes in Computer Science 4702* (2007) 164–175.

# Présentation sur un exemple

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- On dit que 'good' > 'medium' > 'bad'
- On dit que '9' < '10' (« le plus est le mieux »)

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Quels sont ceux qui valent ou qui sont meilleurs que Student\_1 ?
  - En Mathematics : 1,4,5,6
  - En Physics : 1,2,3,4,5,6
  - En Literature : 1,2,3,4,5,6,7,8
  - En Note: 1,2,3,4,5,6,8
  - Globalement sur les 4 critères, on retrouve :  $\{1,4,5,6\} = D_p^+(1)$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Quels sont ceux qui valent ou qui sont meilleurs que Student\_2 ?
  - En Mathematics : 1,2,3,4,5,6
  - En Physics : 1,2,3,4,5,6
  - En Literature : 1,2,3,4,5,6,7,8
  - En Note: 2,3,4,5,6
  - Globalement sur les 4 critères, on retrouve :  $\{2,3,4,5,6\} = D_p^+(2)$



	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

– Au final, sur les 4 critères on a  $D_p^+(x)$  :

« Ensembles de ceux qui valent ou qui sont meilleurs que »

- Student\_1: 1,4,5,6
- Student\_2: 2,3,4,5,6
- Student\_3: 3,4,5,6
- Student\_4: 4,6
- Student\_5: 5
- Student\_6: 6
- Student\_7: 1,2,3,4,5,6,7,8
- Student\_8: 3,4,5,6,8

Ensembles « dominants »

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Quels sont ceux qui valent ou qui sont moins bons que Student\_1 ?
  - En Mathematics : 1,2,3,4,5,6,7,8
  - En Physics : 1,2,3,5,7,8
  - En Literature : 1,2,7
  - En Note: 1,7,8
  - Globalement sur les 4 critères, on retrouve :  $\{1,7\} = D_p^-(1)$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Quels sont ceux qui valent ou qui sont moins bons que Student\_2 ?
  - En Mathematics : 2,3,7,8
  - En Physics : 1,2,3,5,7,8
  - En Literature : 1,2,7
  - En Note: 1,2,7,8
  - Globalement sur les 4 critères, on retrouve :  $\{2,7\} = D_p^-(2)$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Au final, sur les 4 critères on a  $D_p^-(x)$  :

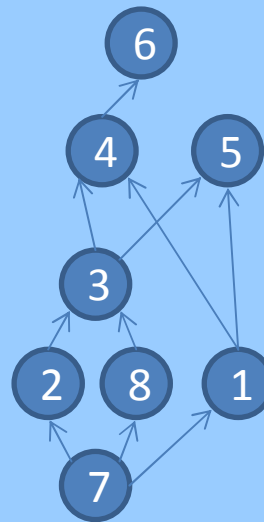
« Ensembles de ceux qui valent ou qui sont moins bons que »

- Student\_1: 1,7
- Student\_2: 2,7
- Student\_3: 2,3,7,8
- Student\_4: 1,2,3,4,7,8
- Student\_5: 1,2,3,5,7,8
- Student\_6: 1,2,3,4,6,7,8
- Student\_7: 7
- Student\_8: 7,8

Ensembles « dominés »

# Ensembles dominants/dominés

- Ensembles dominants (meilleurs que)  $D_p^+(x)$  :
  - Student\_1: 1,4,5,6
  - Student\_2: 2,3,4,5,6
  - Student\_3: 3,4,5,6
  - Student\_4: 4,6
  - Student\_5: 5
  - Student\_6: 6
  - Student\_7: 1,2,3,4,5,6,7,8
  - Student\_8: 3,4,5,6,8



- Ensembles dominés (moins bons que)  $D_p^-(x)$  :
  - Student\_1: 1,7
  - Student\_2: 2,7,8
  - Student\_3: 2,3,7,8
  - Student\_4: 1,2,3,4,7,8
  - Student\_5: 1,2,3,5,7,8
  - Student\_6: 1,2,3,4,6,7,8
  - Student\_7: 7
  - Student\_8: 7,8

	Mathematics	Physics	Literature	Note	OVERALL
Student_1					bad
Student_2					medium
Student_3					medium
Student_4					good
Student_5					good
Student_6					good
Student_7					bad
Student_8					bad

# Approximations

- On s'intéresse à la variable de décision « OVERALL »
- On fait des classes :
  - Au moins « good » = good :  $OVERALL_{\geq \text{good}} = \{4,5,6\}$
  - Au moins « medium » = medium ou good :  $OVERALL_{\geq \text{medium}} = \{2,3,4,5,6\}$
  - Au plus « bad » = bad :  $OVERALL_{\leq \text{bad}} = \{1,7,8\}$
  - Au plus « medium » = medium ou bad :  $OVERALL_{\leq \text{medium}} = \{1,2,6,7,8\}$
 (au moins « bad » et au plus « good » sont inutiles)

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

• Ensembles dominants  $D_p^+(x)$  :  
(ceux qui valent ou qui sont meilleurs que)

- Student\_1: 1,4,5,6
- Student\_2: 2,3,4,5,6
- Student\_3: 3,4,5,6
- Student\_4: 4,6
- Student\_5: 5
- Student\_6: 6
- Student\_7: 1,2,3,4,5,6,7,8
- Student\_8: 3,4,5,6,8

• Ensembles dominés  $D_p^-(x)$  :  
(ceux qui valent ou qui sont moins bons que)

- Student\_1: 1,7
- Student\_2: 2,7,8
- Student\_3: 2,3,7,8
- Student\_4: 1,2,3,4,7,8
- Student\_5: 1,2,3,5,7,8
- Student\_6: 1,2,3,4,6,7,8
- Student\_7: 7
- Student\_8: 7,8

$$\text{OVERALL}_{\geq \text{good}} = \{4,5,6\}$$

$$\text{OVERALL}_{\geq \text{medium}} = \{2,3,4,5,6\}$$

$$\text{OVERALL}_{\leq \text{bad}} = \{1,7,8\}$$

$$\text{OVERALL}_{\leq \text{medium}} = \{1,2,6,7,8\}$$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

• Ensembles dominants  $D_p^+(x)$  :  
(ceux qui valent ou qui sont meilleurs que)

- Student\_1: 1,4,5,6
- Student\_2: 2,3,4,5,6
- Student\_3: 3,4,5,6
- Student\_4: 4,6
- Student\_5: 5
- Student\_6: 6
- Student\_7: 1,2,3,4,5,6,7,8
- Student\_8: 3,4,5,6,8

• Ensembles dominés  $D_p^-(x)$  :  
(ceux qui valent ou qui sont moins bons que)

- Student\_1: 1,7
- Student\_2: 2,7,8
- Student\_3: 2,3,7,8
- Student\_4: 1,2,3,4,7,8
- Student\_5: 1,2,3,5,7,8
- Student\_6: 1,2,3,4,6,7,8
- Student\_7: 7
- Student\_8: 7,8

$OVERALL_{\geq \text{good}} = \{4,5,6\}$   
 $OVERALL_{\geq \text{medium}} = \{2,3,4,5,6\}$   
 $OVERALL_{\leq \text{bad}} = \{1,7,8\}$   
 $OVERALL_{\leq \text{medium}} = \{1,2,6,7,8\}$

Ceux qui ne sont dominés que par des «  $\geq$  good » : 4, 5 et 6  
 $\underline{P}(CI_{\geq \text{good}}) = \{4,5,6\}$  (pour lesquels les dominants sont dans  $\{4,5,6\}$ )

Ceux qui dominent au moins un des «  $\geq$  good » : 4, 5, et 6  
 $\wedge P(CI_{\geq \text{good}}) = \{4,5,6\}$



	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

- Ensembles dominants  $D_p^+(x)$  :  
(ceux qui valent ou qui sont meilleurs que)
  - Student\_1: 1,4,5,6
  - Student\_2: 2,3,4,5,6
  - Student\_3: 3,4,5,6
  - Student\_4: 4,6
  - Student\_5: 5
  - Student\_6: 6
  - Student\_7: 1,2,3,4,5,6,7,8
  - Student\_8: 3,4,5,6,8

- Ensembles dominés  $D_p^-(x)$  :  
(ceux qui valent ou qui sont moins bons que)
  - Student\_1: 1,7
  - Student\_2: 2,7,8
  - Student\_3: 2,3,7,8
  - Student\_4: 1,2,3,4,7,8
  - Student\_5: 1,2,3,5,7,8
  - Student\_6: 1,2,3,4,6,7,8
  - Student\_7: 7
  - Student\_8: 7,8

$OVERALL_{\geq \text{good}} = \{4,5,6\}$   
 $OVERALL_{\geq \text{medium}} = \{2,3,4,5,6\}$   
 $OVERALL_{\leq \text{bad}} = \{1,7,8\}$   
 $OVERALL_{\leq \text{medium}} = \{1,2,6,7,8\}$

Ceux qui ne sont dominés que par des «  $\geq$  medium » : 2, 3, 4, 5 et 6  
 $\underline{P}(CI_{\geq \text{medium}}) = \{2,3,4,5,6\}$  (pour lesquels les dominants sont dans  $\{2,3,4,5,6\}$ )

Ceux qui dominent au moins un des «  $\geq$  medium » : 2, 3, 4, 5, et 6  
 $\wedge P(CI_{\geq \text{medium}}) = \{2,3,4,5,6\}$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Approximations

## $\underline{P}$ et $\wedge P$

• Ensembles dominants  $D_p^+(x)$  :  
(ceux qui valent ou qui sont meilleurs que)

- Student\_1: 1,4,5,6
- Student\_2: 2,3,4,5,6
- Student\_3: 3,4,5,6
- Student\_4: 4,6
- Student\_5: 5
- Student\_6: 6
- Student\_7: 1,2,3,4,5,6,7,8
- Student\_8: 3,4,5,6,8

• Ensembles dominés  $D_p^-(x)$  :  
(ceux qui valent ou qui sont moins bons que)

- Student\_1: 1,7
- Student\_2: 2,7,8
- Student\_3: 2,3,7,8
- Student\_4: 1,2,3,4,7,8
- Student\_5: 1,2,3,5,7,8
- Student\_6: 1,2,3,4,6,7,8
- Student\_7: 7
- Student\_8: 7,8

$OVERALL \geq_{good} = \{4,5,6\}$   
 $OVERALL \geq_{medium} = \{2,3,4,5,6\}$   
 $OVERALL \leq_{bad} = \{1,7,8\}$   
 $OVERALL \leq_{medium} = \{1,2,6,7,8\}$

Ceux qui ne sont dominés que par les «  $\leq_{medium}$  » : 1, 2, 7, 8  
 $\underline{P}(CI_{\leq_{medium}}) = \{1,2,7,8\}$  (les dominés sont dans  $\{1,2,6,7,8\}$ )

Ceux qui sont dominés par au moins un des «  $\leq_{medium}$  » : 1, 2, 3, 4, 6, 7, 8 (un dominant est dans  $\{1,2,6,7,8\}$ )  
 $\wedge P(CI_{\leq_{medium}}) = \{1,2,3,4,6,7,8\}$

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Approximations

## $\underline{P}$ et $\wedge P$

• Ensembles dominants  $D_p^+(x)$  :  
(ceux qui valent ou qui sont meilleurs que)

- Student\_1: 1,4,5,6
- Student\_2: 2,3,4,5,6
- Student\_3: 3,4,5,6
- Student\_4: 4,6
- Student\_5: 5
- Student\_6: 6
- Student\_7: 1,2,3,4,5,6,7,8
- Student\_8: 3,4,5,6,8

• Ensembles dominés  $D_p^-(x)$  :  
(ceux qui valent ou qui sont moins bons que)

- Student\_1: 1,7
- Student\_2: 2,7,8
- Student\_3: 2,3,7,8
- Student\_4: 1,2,3,4,7,8
- Student\_5: 1,2,3,5,7,8
- Student\_6: 1,2,3,4,6,7,8
- Student\_7: 7
- Student\_8: 7,8

$OVERALL \geq_{good} = \{4,5,6\}$   
 $OVERALL \geq_{medium} = \{2,3,4,5,6\}$   
 $OVERALL \leq_{bad} = \{1,7,8\}$   
 $OVERALL \leq_{medium} = \{1,2,6,7,8\}$

Ceux qui ne sont dominés que par les «  $\leq_{bad}$  » : 1, 7, 8  
 $\underline{P}(Cl_{\leq_{bad}}) = \{1,7,8\}$  (les dominés sont dans  $\{1,7,8\}$ )

Ceux qui sont dominés par au moins un des «  $\leq_{bad}$  » : 1, 7, 8  
 $\wedge P(Cl_{\leq_{bad}}) = \{1,7,8\}$

# Règles

- Les règles « certaines » sont générées à partir des approximations inférieures ( $\underline{P}$ )  
« Si  $f_{i1}(x) \geq r_{i1}$  et ... et  $f_{ip}(x) \geq r_{ip}$ , alors  $x \in Cl_t^{\geq}$  »
- Les règles « possibles » sont générés à partir des approximations supérieures ( $\hat{P}$ )  
« X pourrait appartenir à  $Cl_t^{\geq}$  ou la forme: X pourrait appartenir à  $Cl_t^{\leq}$  »

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles

	P
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}

**OVERALL  $\geq$  GOOD**

Mathematics  $\geq$  good : {1,4,5,6}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Mathematics  $\geq$  medium : {1,2,3,4,5,6}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien (inutile a priori)

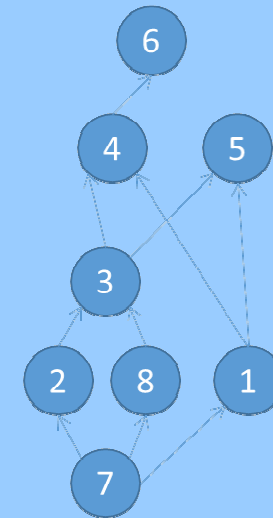
Mathematics  $\leq$  medium : {2,3,7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Mathematics  $\leq$  bad : {7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles

	P
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}



OVERALL  $\geq$  GOOD

Physics  $\geq$  good

: {4,6}  $\subset$  {4,5,6}  $\Rightarrow$  « **Physics  $\geq$  good  $\Rightarrow$  OVERALL  $\geq$  good** »

Physics  $\geq$  medium

: {1,2,3,4,5,6}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Physics  $\leq$  medium

: {1,2,3,5,7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Physics  $\leq$  bad

: {7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles

	P
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}

OVERALL  $\geq$  GOOD

Literature  $\geq$  good : {5,6}  $\subset$  {4,5,6}  $\Rightarrow$  « **Literature  $\geq$  good  $\Rightarrow$  OVERALL  $\geq$  good** »

Literature  $\geq$  medium : {3,4,5,6,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Literature  $\leq$  medium : {1,2,3,4,7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

Literature  $\leq$  bad : {1,2,7}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles

	P
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}

## OVERALL $\geq$ GOOD

- Note  $\geq 16$  : {5}  $\subset$  {4,5,6}  $\Rightarrow$  « **Note  $\geq 16 \Rightarrow$  OVERALL  $\geq$  good** »
- Note  $\geq 15$  : {5,6}  $\subset$  {4,5,6}  $\Rightarrow$  « **Note  $\geq 15 \Rightarrow$  OVERALL  $\geq$  good** »
- Note  $\geq 12$  : {4,5,6}  $\subset$  {4,5,6}  $\Rightarrow$  « **Note  $\geq 12 \Rightarrow$  OVERALL  $\geq$  good** »
- Note  $\geq 11$  : {3,4,5,6}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien
- Note  $\leq 7$  : {7}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien
- Note  $\leq 9$  : {1,7,8}  $\not\subset$  {4,5,6}  $\Rightarrow$  rien, etc.



	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles multiples

	<u>P</u>
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}

OVERALL  $\geq$  GOOD

Mathematics  $\geq$  good et Note  $\geq$  12 : {4,5,6}  $\subset$  {4,5,6}

$\Rightarrow$  « **Mathematics  $\geq$  good et Note  $\geq$  12  $\Rightarrow$  OVERALL  $\geq$  good** »

OVERALL  $\geq$  GOOD

Mathematics  $\geq$  good et Literature  $\geq$  good : {5,6}  $\subset$  {4,5,6}

$\Rightarrow$  « **Mathematics  $\geq$  medium et Literature  $\geq$  good  $\Rightarrow$  OVERALL  $\geq$  good** »

	Mathematics	Physics	Literature	Note	OVERALL
Student_1	good	medium	bad	9	bad
Student_2	medium	medium	bad	10	medium
Student_3	medium	medium	medium	11	medium
Student_4	good	good	medium	12	good
Student_5	good	medium	good	16	good
Student_6	good	good	good	15	good
Student_7	bad	bad	bad	7	bad
Student_8	bad	bad	medium	9	bad

# Règles multiples

	P
OVERALL $\geq$ good	Ceux qui ne sont dominés que par des « $\geq$ good » : {4,5,6}
OVERALL $\geq$ medium	Ceux qui ne sont dominés que par des « $\geq$ medium » : {2,3,4,5,6}
OVERALL $\leq$ medium	Ceux qui ne sont dominés que par les « $\leq$ medium » : {1,2,7,8}
OVERALL $\leq$ bad	Ceux qui ne sont dominés que par les « $\leq$ bad » : {1,7,8}

OVERALL  $\geq$  MEDIUM

Mathematics  $\geq$  medium et Literature  $\geq$  medium : {3,4,5,6}  $\subset$  {2,3,4,5,6}

$\Rightarrow$  « **Mathematics  $\geq$  medium et Literature  $\geq$  medium  $\Rightarrow$  OVERALL  $\geq$  medium** »

OVERALL  $\geq$  MEDIUM

Mathematics  $\geq$  medium et Note  $\geq$  10 : {2,3,4,5,6}  $\subset$  {2,3,4,5,6}

$\Rightarrow$  « **Mathematics  $\geq$  medium et Note  $\geq$  10  $\Rightarrow$  OVERALL  $\geq$  medium** »

# **Fin - Rough Set Approach to Knowledge Discovery about Preferences**

# Objectifs

Comparer 2 approches DRSA et OPLS-DA pour analyser les données métabolomiques

- pour évaluer la discrimination entre les sujets SLA et contrôles
- pour identifier et valider des biomarqueurs diagnostiques

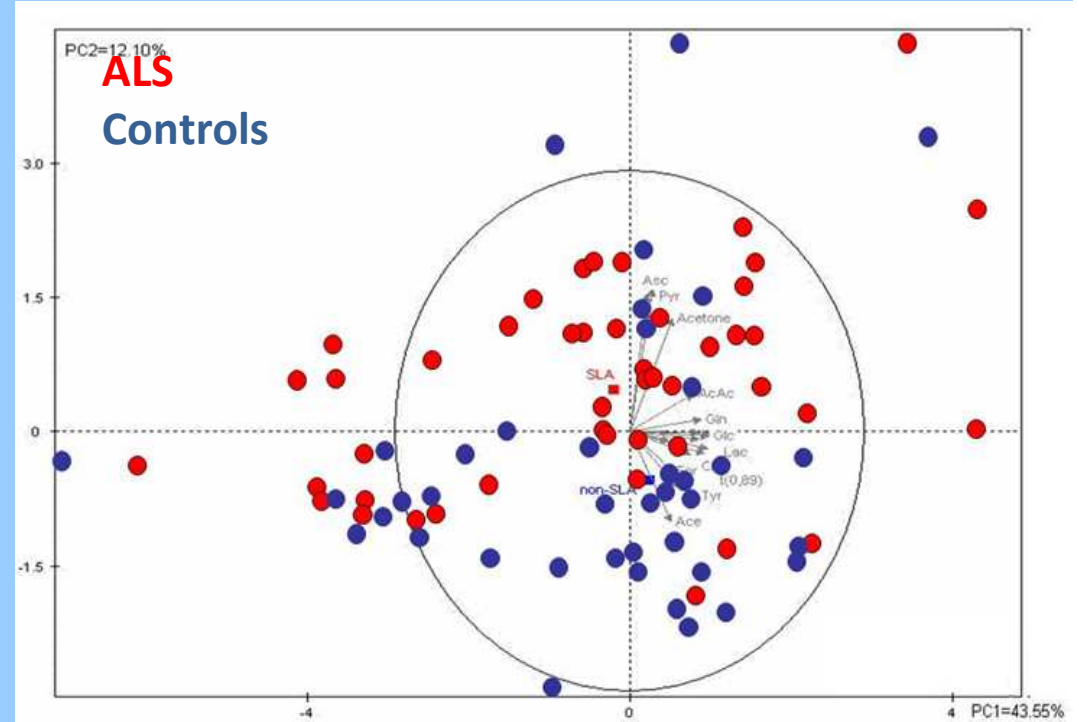
# Etude initiale : métabolomique ciblée

-Analyse du métabolome du LCR de patients SLA et contrôles par spectroscopie par résonance magnétique

## Analyse univariée (p: 0,003)

Metabolites ( $\mu\text{mol/L}$ )	ALS (n = 50)	Non ALS (n = 44)	p
AHBT	61 (0-137)	59 (0-177)	0.17
Ethanol	24 (0-170)	24 (0-460)	0.29
Alanine	56 (6-596)	61 (0-214)	0.74
Acetate	53 (9-389)	104 (2-366)	<b>0.0002</b>
BHBT	17 (4-85)	27 (0-134)	0.97
Lactate	2076 (444-3892)	2020 (94-5322)	0.81
Acetoacetate	9 (0-29)	8 (0-29)	0.89
Acetone	11 (0-59)	5 (0-23)	0.015
Glutamine	716 (139-1272)	785 (35-1296)	0.24
Pyruvate	68 (3-128)	17 (0-201)	<b>0.002</b>
Citrate	283 (22-49)	285 (12-754)	0.51
Creatine/creatinine	155 (25-316)	160 (9-450)	0.65
Glucose	3240 (560-7166)	3174 (150-8928)	0.98
Fructose	428 (0-1224)	419 (0-967)	0.82
Ascorbate	25 (0-152)	0 (0-165)	<b>0.003</b>
Tyrosine	12 (0-48)	8 (0-117)	0.35
Formate	28 (3-91)	30 (1-187)	0.35

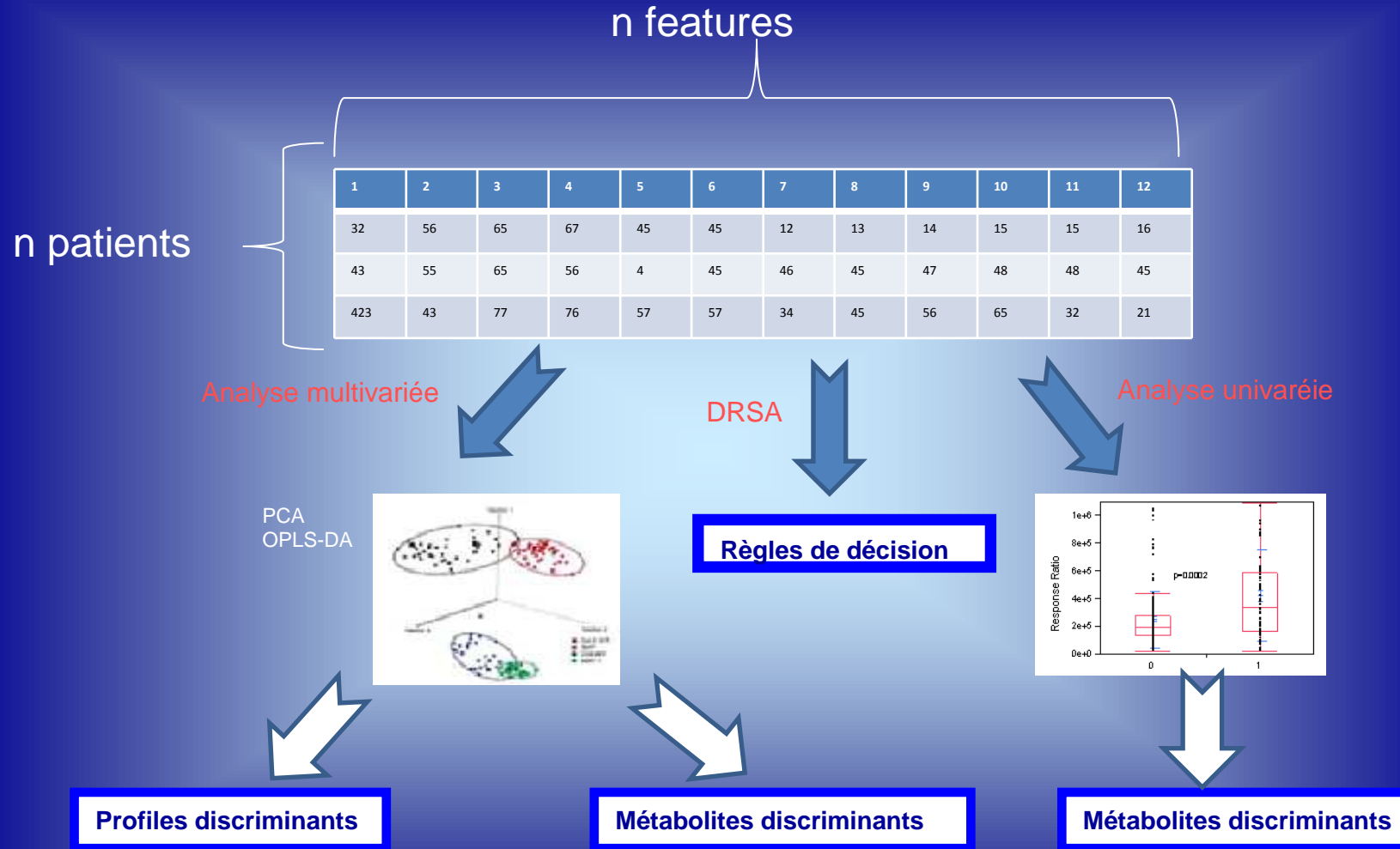
## Analyse multivariée (PCA)



⇒ Perturbation du métabolisme du glucose

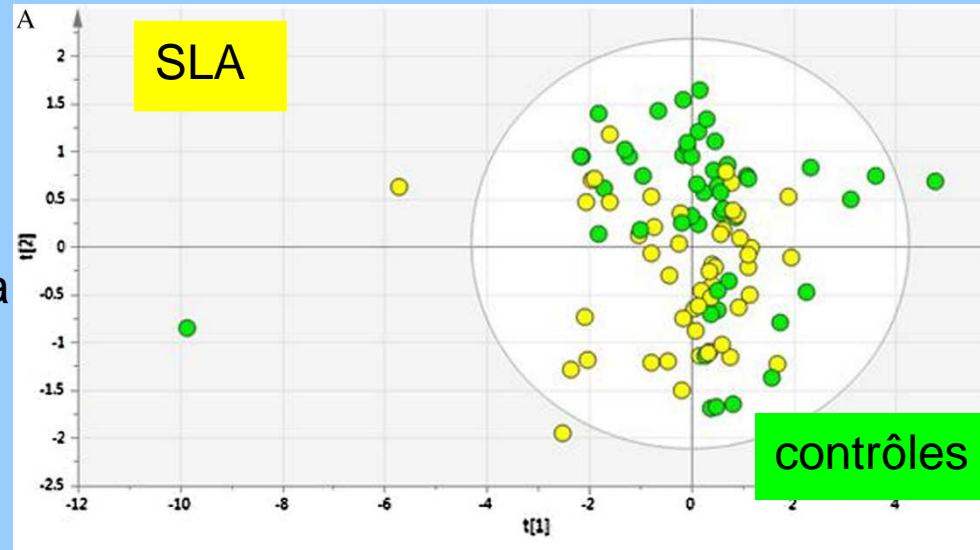
⇒ 81.6% de prédiction correcte

# Ré-analyse des données

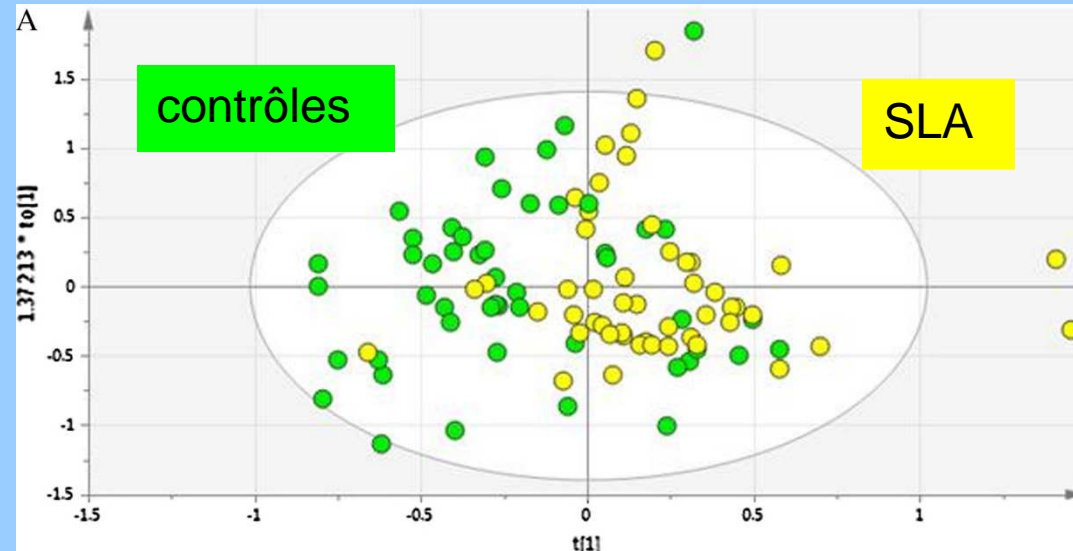


# Analyse des données PCA

Explication de 64,1% de la variation des métabolites



# Analyse des données OPLS-DA



$R^2X(\text{cum}) = 0.999$ ,  $R^2Y(\text{cum}) = 0.245$ , and  $Q^2(\text{cum}) = 0.167$



# Analyse des données OPLS-DA

-à partir de 17 métabolites :

Sensibilité : 78%

Spécificité : 67,3%

VPP: 70,9%

VPN: 75%

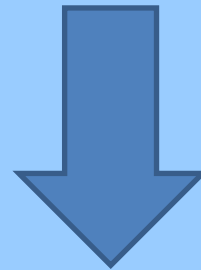
-à partir de 4 métabolites :

Sensibilité : 72%

Spécificité : 82%

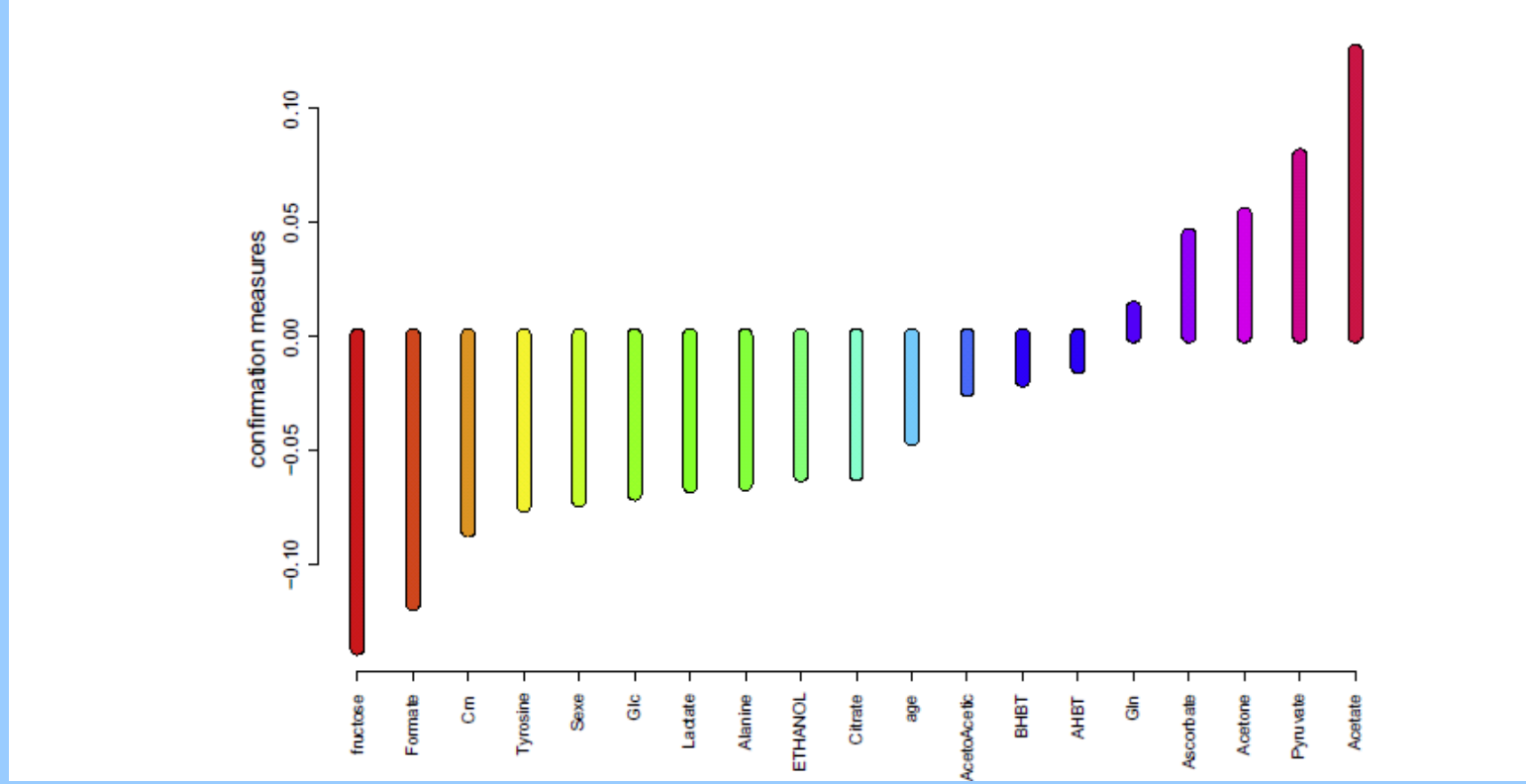
VPP: 80%

VPN: 74,1%



Courbe ROC : **AUC=0,79**  
avec une sensibilité de 90%  
et spécificité à 69,5%

# DRSA



# Résultats de l'approche Rough Set Approach to Knowledge Discovery about Preferences

- The rules induced from the information table structured using DRSA represent relevant patterns of condition–decision relationships, which are free of inessential and redundant information. Some of the most relevant of all the rules that constitute the DRSA model are as follows.
  1. if AcetateP  $\geq 0.125$  and Ascorbate  $\leq 0.01964669$  then patient is not ALS;
    - {strength: 0.21, confirmation s = 0.57}
  2. if GlnP  $\geq 0.804629579$  and Acetone  $\leq 0.007626821$  then patient is not ALS;
    - {strength: 0.11, confirmation s = 0.37}
  3. if Pyruvate  $\leq 0.029593053$  and CrnP  $\geq 0.192354842$  then patient is not ALS;
    - {strength: 0.09, confirmation s = 0.2}
  4. if Acetate  $\leq 0.055457394$  and Ascorbate  $\leq 0.0786$  then patient is ALS;
    - {strength: 0.25, confirmation s = 0.55}
  5. if Acetone  $\in [0.0087; 0.009534583]$  then patient is ALS;
    - {strength: 0.04, confirmation s = 0.5}
  6. if Alanine  $\leq 0.039024531$  and PyruvateP  $\geq 0.0326$  then patient is ALS;
    - {strength: 0.08, confirmation s = 0.55}

# Discussion

- Paramètres de validation interne des modèles différents mais résultats idem pour les
  - Performances de prédictions
  - Métabolites discriminants
    - Acétate + pyruvate
    - Alanine + pyruvate
    - Acétate + ascorbate
  - métabolisme du glucose
- Paramètre identifié par DRSA : béta-hydroxybutyrate→compatible avec acétate et acétone

# Discussion

- Prise en compte de la nature et du sens de variation des métabolites selon les métabolites associés
- Confirmation Bayésienne → acétate, pyruvate, acétone, **ascorbate**

# Conclusion

- DRSA: méthode complémentaire utilisable pour analyser la métabolomique
- DRSA pertinente pour faire des hypothèses mécanistiques des voies physiopathologiques
- Relation condition-décision : prometteur pour les prédictions dans le domaine médical

# Perspectives

- Intégrer d'autres types de données
  - Données démographiques
  - Données cliniques
  - Données paracliniques
- Prédire des paramètres plus subtiles : évolution de la maladie

# Remerciements



## TOURS

**U930** (équipe 2)

C. Andres

P. Corcia

P. Vourc'h

S. Mavel

C. Veyrat-Durebex

C. Antar

**PPF**

P. Emond

C. Bocca

L. Nadal-Desbarats

**CNRS LI EA**

**6300, OC ERL**

**CNRS 6305**

JC-Billaut

## PARIS

PF. Pradat

## LILLE

D. Devos

C. Moreau

## Pologne

Institute of Computing Science

J Blaszczynski

R Slowinski