

## Apprentissage multi-label

Haytham Elghazel

Laboratoire d'InfoRmatique en Image et Systèmes d'information

*Pôle Data Science, Equipe DM2L*



**INSA**



UNIVERSITÉ  
LUMIÈRE  
LYON 2



# Apprentissage multi-label

	$X_1$					$X_D$	$M$	$Sa$	$So$	$A$	$C$	$N$
$E_1$	a	b	...	c	d		1	1	1	1	0	0
$E_1$	e	f	...	g	h		0	1	0	1	1	0
$E_1$	i	j	...	k	l		0	0	1	0	0	1

**M:** Mer

**Sa:** Sable

**So:** Soleil

**A:** Arbre

**C:** Chameau

**N:** Neige

## Fonction de Classification

$$f : \mathcal{X} \longrightarrow \{0, 1\}^L$$

### Domaines d'application:

Catégorisation de texte,

Classification des fonctions génétiques,

Annotation sémantique des images, vidéos.



$E_1$  : Image 1



$E_2$  : Image 2



$E_3$  : Image 3

# Apprentissage multi-label

## ■ Représentation des données Multi-label

	$X_1$	$X_2$	$\dots$	$X_D$	$\lambda_1$	$\lambda_2$	$\dots$	$\lambda_Q$
$E_1$	$x_{1,1}$	$x_{1,2}$	$\dots$	$x_{1,D}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\dots$	$\lambda_{1,Q}$
$E_2$	$x_{2,1}$	$x_{2,2}$	$\dots$	$x_{2,D}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\dots$	$\lambda_{2,Q}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$E_N$	$x_{N,1}$	$x_{N,2}$	$\dots$	$x_{N,D}$	$\lambda_{N,1}$	$\lambda_{N,2}$	$\dots$	$\lambda_{N,Q}$

$D$  : Le Nombre de variables;  $N$  : Le nombre d'instances;  $Q$  : Le nombre de labels.

## ■ La fonction de Classification

$$f : \mathcal{X} \longrightarrow \{0, 1\}^Q$$

# Métriques d'évaluation

## ■ Hamming Loss :

- ❑ Elle mesure le taux d'erreurs par étiquette, en prenant en compte les faux positifs (étiquettes incorrectement prédites) et les faux négatifs (étiquettes manquées).

$$\text{Hamming Loss} = \frac{1}{N \times Q} \sum_{i=1}^N \sum_{j=1}^Q 1(y_{ij} \neq \hat{y}_{ij})$$

- ❑ où **N** est le nombre d'exemples, **Q** le nombre d'étiquettes, **y<sub>ij</sub>** l'étiquette réelle et **ŷ<sub>ij</sub>** l'étiquette prédite.

## ■ 0/1 loss (Subset Accuracy)

- ❑ Elle mesure si l'ensemble des étiquettes prédites pour un exemple correspond exactement à l'ensemble des étiquettes réelles. Elle est plus stricte que la Hamming Loss.

$$0/1 \text{ Loss} = \frac{1}{N} \sum_{i=1}^N 1(Y_i \neq \hat{Y}_i)$$

- ❑ où **Y<sub>i</sub>** est l'ensemble des étiquettes réelles pour un exemple, et **Ŷ<sub>i</sub>** est l'ensemble des étiquettes prédites.

# Métriques d'évaluation

- ❑ Ces deux métriques fournissent des perspectives différentes sur les performances d'un modèle multi-label.

Individu	Étiquettes réelles	Prédictions
1	[1, 0, 1]	[1, 0, 0]
2	[0, 1, 1]	[0, 1, 1]
3	[1, 1, 0]	[1, 0, 0]

$$\text{Hamming Loss} = \frac{1}{N \times Q} \sum_{i=1}^N \sum_{j=1}^Q 1(y_{ij} \neq \hat{y}_{ij}) = \frac{1}{3 \times 3} \times (1 + 0 + 2) = \frac{3}{9} = 0,3333$$

$$0/1 \text{ Loss} = \frac{1}{N} \sum_{i=1}^N 1(Y_i \neq \hat{Y}_i) = \frac{1}{3} \times (1 + 0 + 1) = \frac{2}{3} = 0,6667$$

# Métriques d'évaluation

- Le **F1-Score** est une métrique utilisée pour évaluer la performance d'un modèle. Il existe deux variantes principales pour calculer le F1-Score dans ce contexte Multi label : **Micro F1** et **Macro F1**.
- Elles permettent d'obtenir des perspectives différentes sur les performances d'un modèle, en particulier lorsque le jeu de données est déséquilibré.
- **Micro F1** : Elle prend en compte toutes les étiquettes ensemble comme si elles appartenaient à une seule classe. Le F1-Score est calculé en agrégeant les vrais positifs (TP), faux positifs (FP) et faux négatifs (FN) sur toutes les étiquettes avant d'effectuer le calcul. Cette métrique est particulièrement utile lorsque vous voulez accorder plus de poids aux classes fréquentes.

$$\text{Micro F1} = \frac{2 \times \text{Micro Précision} \times \text{Micro Rappel}}{\text{Micro Précision} + \text{Micro Rappel}}$$

$$\begin{aligned}\text{Micro Précision} &= \frac{\sum TP}{\sum (TP + FP)} \\ \text{Micro Rappel} &= \frac{\sum TP}{\sum (TP + FN)}\end{aligned}$$

# Métriques d'évaluation

- ❑ **Macro F1** : Elle traite chaque étiquette indépendamment. Elle calcule le F1-Score pour chaque étiquette individuellement, puis prend la moyenne arithmétique de ces scores. Cette approche donne un poids égal à chaque étiquette, quel que soit le nombre d'occurrences de cette étiquette dans le jeu de données.

$$\text{Macro F1} = \frac{1}{Q} \sum_{j=1}^Q \text{F1-Score}_j$$

# Algorithmes

## ■ Apprentissage multi-label

- Méthodes d'adaptation
- Méthodes de Transformation
- Méthodes ensemblistes



# Méthodes d'adaptation



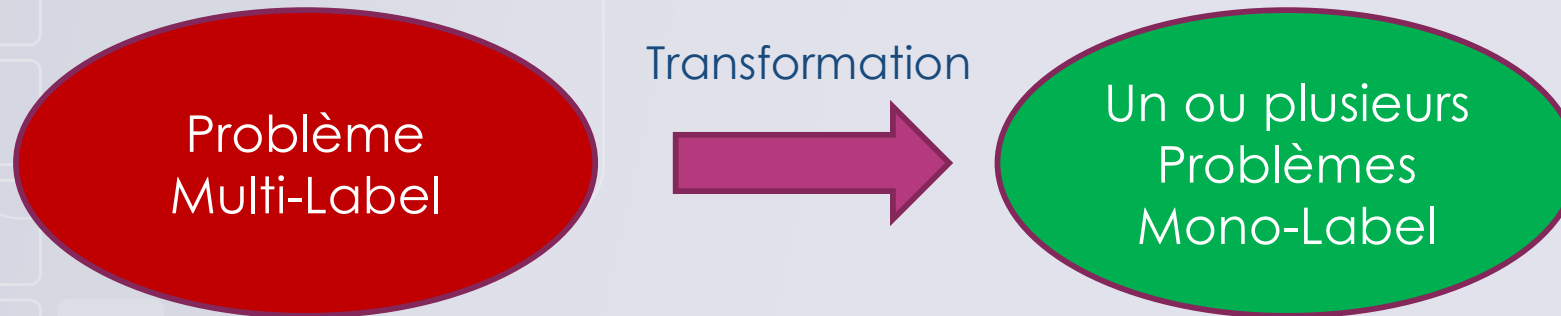
Méthodes adaptant les algorithmes de classification existants:

- $k$  plus proches voisins (**MLkNN**)
- Arbres de décision (**PCT**)

Predictive **C**lustering **T**ree (PCT)

$$\bar{Var}(E) = \sum_{i=1}^Q Gini(E, \mathcal{Y}_i), Gini(E, \mathcal{Y}_i) = 1 - \sum_{j=1}^{C_i} p_{c_{ij}}^2$$

# Méthodes de transformation



- a) **B**inary **R**elevance methods (**BR**),
- b) Pair-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **Binary Relevance methods (BR)**,
- b) Pair-wise methods,
- c) **Label Power-set methods (LP)**.

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

a) **Binary Relevance methods (BR)**,

b) Pair-wise methods,

c) **Label Power-set methods (LP)**.

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **Binary Relevance methods (BR)**,
- b) Pair-wise methods,
- c) **Label Power-set methods (LP)**.

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

**BR**  
Transformation  


# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) Pair-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

**BR**  
Transformation  


$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$
a	b	c	d	1
e	f	g	h	0
i	j	k	l	1

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) Pair-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

**BR**  
Transformation  


$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	
a	$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_2$
e	a	b	c	d	1
i	e	f	g	h	0
	i	j	k	l	1

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **Binary Relevance methods (BR)**,
- b) Pair-wise methods,
- c) **Label Power-set methods (LP)**.

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

**BR**  
Transformation



$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$			
a	$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_2$		
e	a	$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_3$	
i	e	a	b	c	d	1	
	i	e	f	g	h	1	
		i	j	k	l	1	



# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **Binary Relevance methods (BR)**,
- b) Pair-wise methods,
- c) **Label Power-set methods (LP)**.

$x_1$	$x_2$	$x_3$	$x_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1

**BR**  
Transformation



$x_1$	$x_2$	$x_3$	$x_p$	$\lambda_1$		
a	$x_1$	$x_2$	$x_3$	$x_p$	$\lambda_2$	
e	a	$x_1$	$x_2$	$x_3$	$x_p$	$\lambda_3$
i	e	a	b	c	d	1
	i	e	f	g	h	1
		i	j	k	l	1

$$h_j : \mathcal{X} \longrightarrow \{0, 1\}$$

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

# Méthodes de transformation

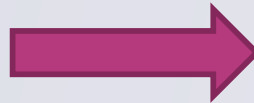
## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
m	n	o	p	1	1	0

Pair-wise  
Transformation



$$\text{Pour } \lambda_i \neq \lambda_j ; \lambda'_{i,j} = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{si } \lambda_j = 1 \end{cases}$$

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
m	n	o	p	1	1	0

Pair-wise  
Transformation



$$\text{Pour } \lambda_i \neq \lambda_j ; \lambda'_{i,j} = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{si } \lambda_j = 1 \end{cases}$$

$h_{1,2}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,2}$
a	b	c	d	1
i	j	k	l	0

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
m	n	o	p	1	1	0

Pair-wise  
Transformation



$$\text{Pour } \lambda_i \neq \lambda_j ; \lambda'_{i,j} = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{si } \lambda_j = 1 \end{cases}$$

$h_{1,2}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,2}$
a	b	c	d	1
i	j	k	l	0

$h_{1,3}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,3}$
i	j	k	l	0
m	n	o	p	1

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
m	n	o	p	1	1	0

Pair-wise  
Transformation



$$\text{Pour } \lambda_i \neq \lambda_j ; \lambda'_{i,j} = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{si } \lambda_j = 1 \end{cases}$$

$h_{1,2}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,2}$
a	b	c	d	1
i	j	k	l	0

$h_{1,3}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,3}$
i	j	k	l	0
m	n	o	p	1

$h_{2,3}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{2,3}$
a	b	c	d	0
m	n	o	p	1

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) **P**air-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
m	n	o	p	1	1	0

Pair-wise  
Transformation



Pour  $\lambda_i \neq \lambda_j$  ;  $\lambda'_{i,j} = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{si } \lambda_j = 1 \end{cases}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,2}$
a	b	c	d	1
i	j	k	l	0

$h_{1,3}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{1,3}$
i	j	k	l	0
m	n	o	p	1

$h_{2,3}$

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda'_{2,3}$
a	b	c	d	0
m	n	o	p	1

$$\frac{Q(Q-1)}{2}$$

# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) Pair-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
f	g	c	d	1	0	1
j	k	g	h	0	1	1

**LP**  
Transformation  


$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_{1,2,3}$
a	b	c	d	1
e	f	g	h	2
i	j	k	l	3
f	g	c	d	1
j	k	g	h	3



# Méthodes de transformation

## A- Apprentissage Multi-Label

b) Méthodes de Transformation:

- a) **B**inary **R**elevance methods (**BR**),
- b) Pair-wise methods,
- c) **L**abel **P**ower-set methods (**LP**).

$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_1$	$\lambda_2$	$\lambda_3$
a	b	c	d	1	0	1
e	f	g	h	1	1	1
i	j	k	l	0	1	1
f	g	c	d	1	0	1
j	k	g	h	0	1	1

**LP**  
Transformation  


$X_1$	$X_2$	$X_3$	$X_4$	$\lambda_{1,2,3}$
a	b	c	d	1
e	f	g	h	2
i	j	k	l	3
f	g	c	d	1
j	k	g	h	3

$h'$

$$h' : \mathcal{X} \longrightarrow P(\mathcal{L})$$

Avec  $P(\mathcal{L})$  les Power sets de  $\mathcal{L}$

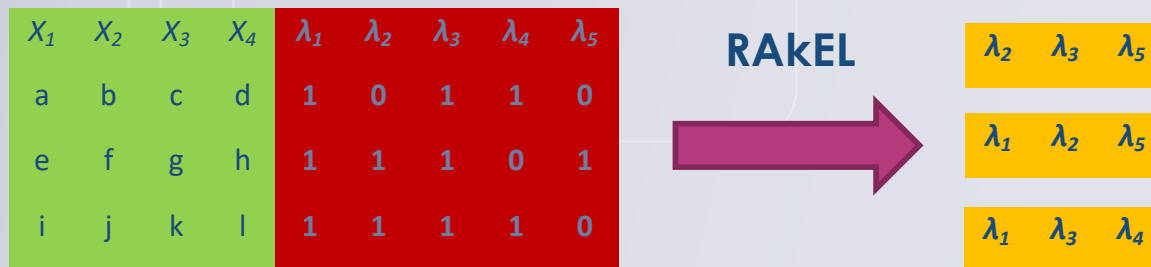
# Méthodes ensemblistes

a) **R**andom **F**orest **P**redictive **C**lustering **T**ree (**RFPCT**)

b) **R**andom **k**-lab**EL**sets (**RAkEL**)

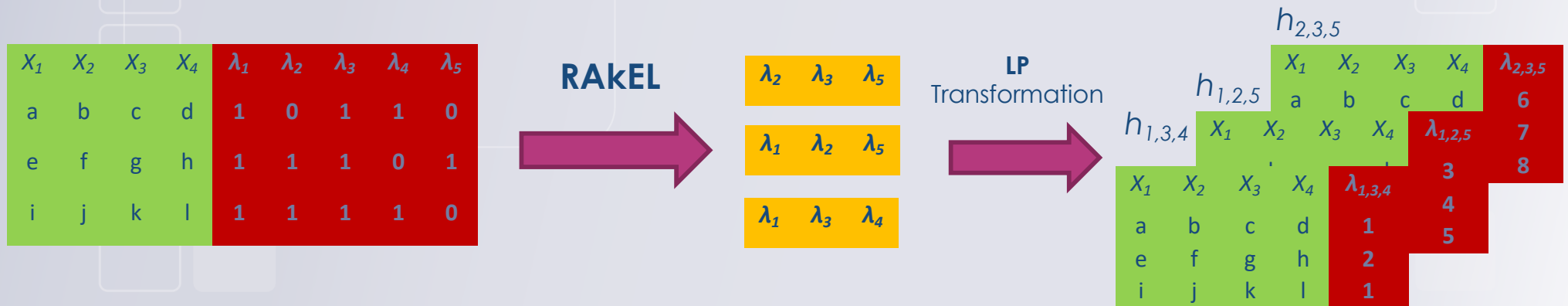
# Méthodes ensemblistes

- a) Random Forest Predictive Clustering Tree (RFPCT)
- b) Random k-labelsets (RAkEL)



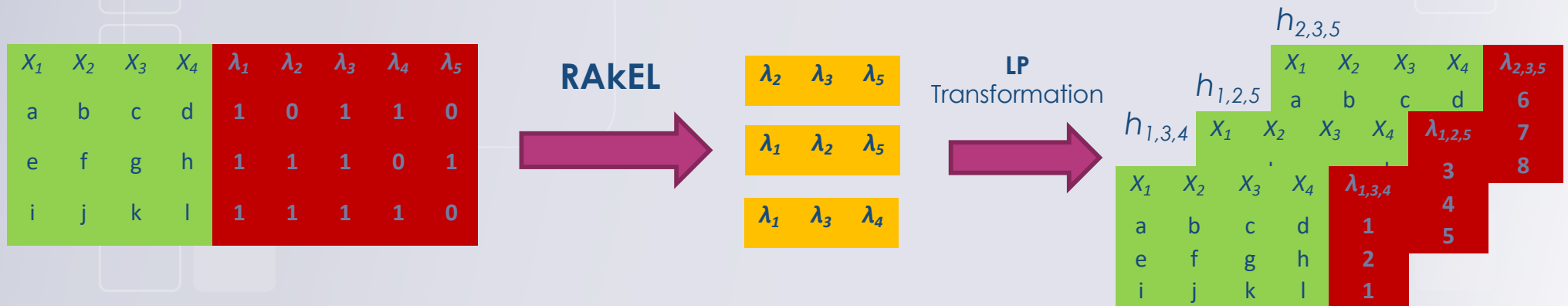
# Méthodes ensemblistes

- a) Random Forest Predictive Clustering Tree (RFPCT)
- b) Random k-labelsets (RAkEL)



# Méthodes ensemblistes

- a) Random Forest Predictive Clustering Tree (RFPCT)
- b) Random k-labelsets (RAkEL)



- c) Ensemble Classifier Chain (ECC)