

Concours au Poste Professeur.e Assistant Monge en Apprentissage Statistique Équipe SIMPAS, CMAP École Polytechnique

Mokhtar Z. Alaya

Équipe APP - LITIS, Université Rouen Normandie

<https://mzalaya.github.io/>

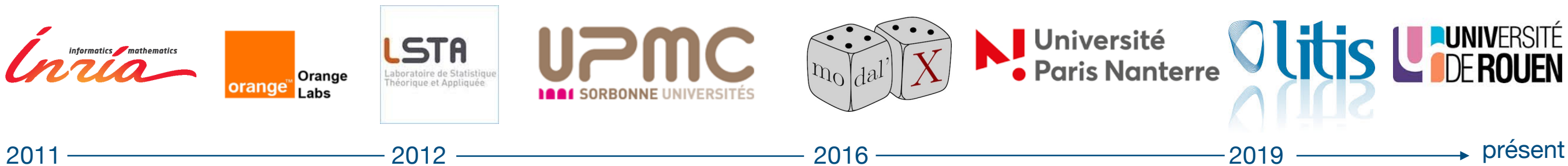


12 Juin 2020

Plan

1. Parcours
2. Enseignement à ℓ^X
3. Résultat de Recherche Marquant
4. Intégration dans l'Équipe SIMPAS

Parcours



- **2010 – 2011: Master 2 Probabilités et Modèles Aléatoires**, Université Pierre et Marie Curie
Stage de recherche **M2**, INRIA-ENS et Orange Labs
- **2011 – 2012: Master 2 Statistique**, Université Pierre et Marie Curie
Stage de recherche **M2**, LSTA, Université Pierre et Marie Curie
- **2012 – 2016: Doctorant-Moniteur**, LSTA, Université Pierre et Marie Curie
Lauréat Allocation de l'École Doctorale ED 386
Direction: Stéphane Gaïffas et Agathe Guilloux
Segmentation de Processus de Comptage et Modèles Dynamiques
- **2015 – 2016: ATER**, LSTA, Université Pierre et Marie Curie
- **2016 – 2017: ATER**, Modal'X, Université Paris Nanterre
- **2017 – Oct 2018: Postdoc 1**, Modal'X, Université Paris Nanterre
Lauréat Postdoc Fondation Sciences Mathématiques de Paris
- **Jan 2019 – présent : Postdoc 2**, Équipe APP, LITIS, Université Rouen Normandie,
Projet ANR OATMIL (OptimAI Transport for Machine Learning)



Bilan d'Enseignements Précédents

Unité d'Enseignement	Formation	Charge horaire	Établissement
Cours Statistique en Grande Dimension	Séminaire Doctorants-Aters Modal'X	10h	Université Paris Nanterre
TD Statistiques	Licence 2 Psychologie	40h	Université Paris Nanterre
TD Méthodologie de la Mesure en Sciences Humaines	Licence 1 Psychologie	60h	Université Paris Nanterre
TD Mathématiques Générales	Licence 1 Économie et Gestion	48h	Université Paris Nanterre
TD Série Temporelles	Master 1	24h	Institut de Statistique - Univ. Pierre et Marie Curie
TD Analyse de Données et Régression	Licence 3 Mathématiques	72h	Univ. Pierre et Marie Curie
TD de Modèle Linéaire Multivarié	Master 1	24h	Institut de Statistique - Univ. Pierre et Marie Curie
TD Algèbre et Géométrie	Licence 2 Mathématiques	72h	Univ. Pierre et Marie Curie
Cours Probabilités et Statistiques	3ème année Génie Mécanique	16h	Polytech Univ. Pierre et Marie Curie

Enseignements à l'*X*

Mes Compétences

Expériences d'enseignement théorique et appliqué

Expérience d'enseignement à différents niveaux et types d'auditoires

Possibilité d'intervenir dans toutes filières de formation concernées

Ingénieur 1A / Bachelor 3

MAP361
Aléatoire

MAP361P
Python

Ingénieur 2A

MAP433
Statistiques

Ingénieur 3A / Master 1

MAP532
Math. Found. DS

MAP541
Advanced ML

MAP531
Stats with R

MAP569
ML II

MAP534
ML I

MAP553
Regression,
Theory + App

MAP535
Regression

MAP545
Optim / DL

MAP536
Python for DS

Master 2

MAP630
Refresh. Stat.

MAP667
Intro. ML

MAP667K
Stat. GD

MAP652B
Elements Stat.

Executive Master

Binarsity: A Penalization for One-Hot Encoded Features in Linear Supervised Learning
JMRL'19

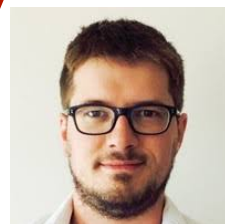
Collective Matrix Completion
JMLR'19

Binacox: Automatic Cut-Points Detection in High-Dimensions Cox Model
soumis à Biometrika'20

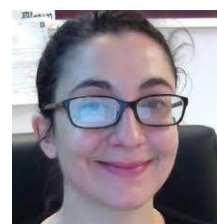
High-Dimensional Time-Varying Aalen and Cox Models
en révision dans J. Nonparam. Stat.'20

Learning the intensity of Time Events with Change Points
IEEE. Trans. Info.'15

Statistique en Grande Dimension



Pr. S. Gaïffas



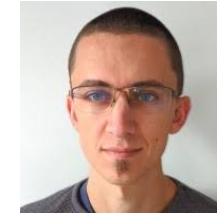
Pr. A. Guillaoux



Dr. O. Klopp



Dr. S. Bussy



Dr. T. Allart



Dr. S. Lemler

..... Co-Auteurs



Pr. A. Rakotomamonjy



Pr. G. Gasso



Dr. M. Bérar



Dr. L. Chapel



Dr. R. Hérault

Machine Learning avec du Transport Optimal

Screening Sinkhorn Algorithm for Regularized Optimal Transport
NeurIPS'19

Non-aligned Distribution using Metric Measure Embedding and Optimal Transport
soumis à NeurIPS'20

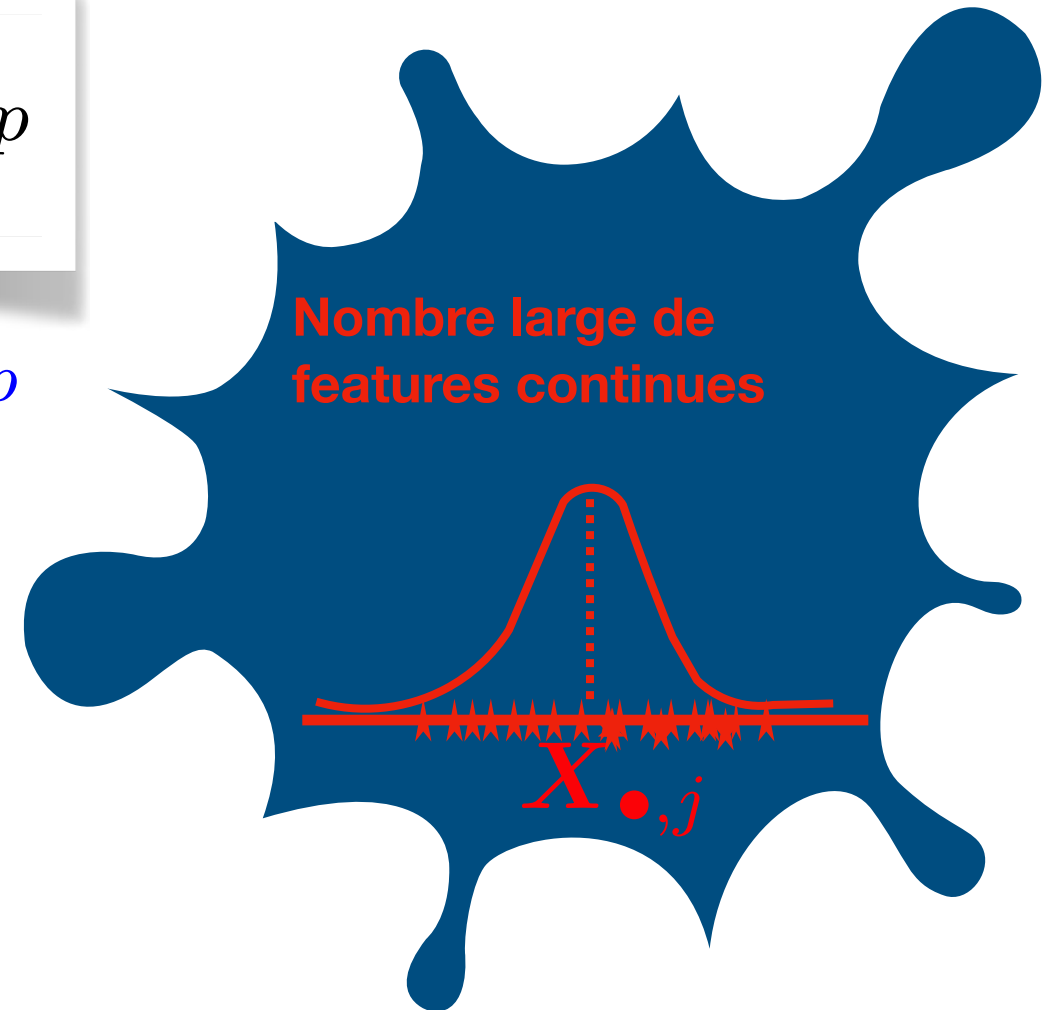
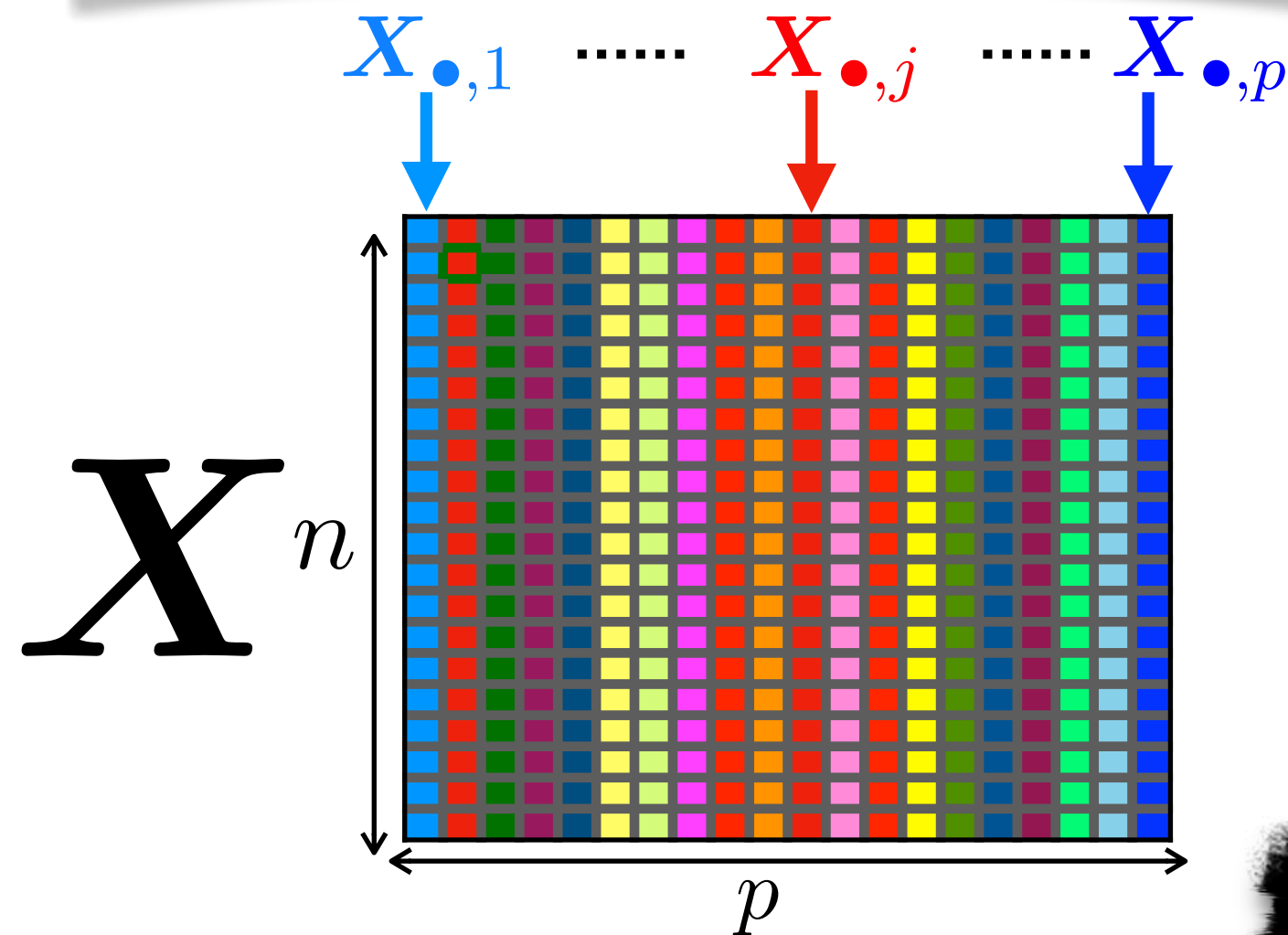
Partial Optimal Transport with Applications on Positive-Unlabeled Learning
soumis à NeurIPS'20

Open Set Domain Adaptation using Optimal Transport
en révision dans ECML-PKDD'20

Production Scientifique

Matrice de features

$$\mathbf{X} = [x_{i,j}]_{1 \leq i \leq n; 1 \leq j \leq p} \in \mathbb{R}^{n \times p}$$



Encodage One-Hot

(Lieu et al., 2002)

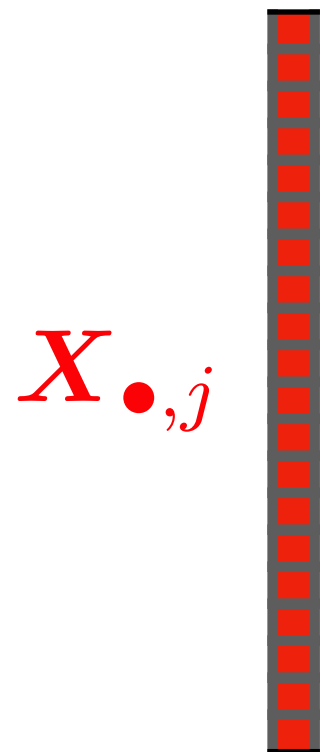
(Wu and Coggeshall, 2012)

Pre-processing de binarisation

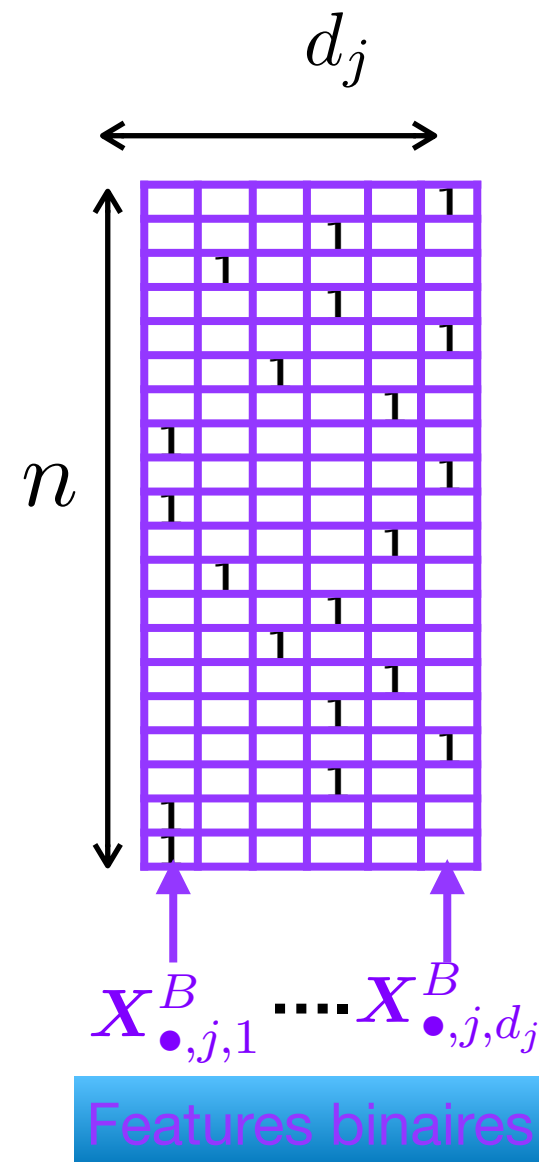
Chaque j-feature $X_{\bullet,j}$ est remplacée par $d_j \geq 2$ features binaires contenant que des zeros et des uns

$$X_{\bullet,j,1}^B, \dots, X_{\bullet,j,d_j}^B$$

Partition de l'ensemble des valeurs de $X_{\bullet,j}$ sur des intervalles $I_{j,1}, \dots, I_{j,d_j}$



$$x_{i,j,k}^B = \begin{cases} 1, & \text{si } x_{i,j} \in I_{j,k}, \\ 0, & \text{sinon} \end{cases}$$



Binarsity : prédiction à grande échelle via la sparsité induite par la binarisation de variables

The diagram illustrates the binarisation process of a matrix X into a binary matrix XB . The matrix XB is partitioned into five blocks of widths $d_1 = 4$, $d_2 = 6$, $d_3 = 6$, $d_4 = 9$, and $d_5 = 10$, with a total height of n . The blocks are color-coded: red, green, blue, orange, and light blue. The diagram shows the transformation of matrix X into matrix XB via a binarisation process, resulting in a matrix with a block structure defined by widths d_1 through d_5 .

1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0



$$d = \sum_{j=1}^p d_j$$

$$\sum_{j=1}^n d_j I_{j,1} = \left[q_j(0), q_j\left(\frac{1}{d_j}\right) \right] \text{ and } I_{j,k} = \left(q_j\left(\frac{k-1}{d_j}\right), q_j\left(\frac{k}{d_j}\right) \right]$$

Coefficients de l'encodage one-hot

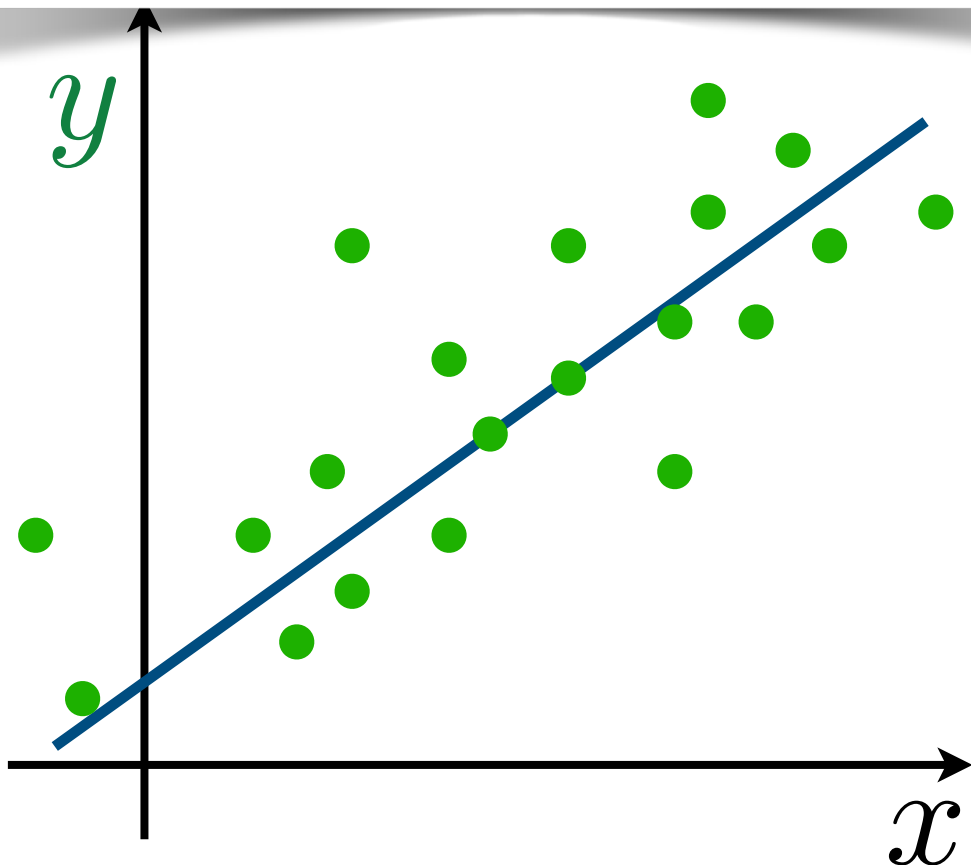
Pour chaque feature binaire $X_{\bullet,j,k}^B$ correspond un paramètre $\theta_{j,k}$.

Le paramètre de binarisation associé à j -th feature est noté

$$\theta_{j,\bullet} = [\theta_{j,1} \cdots \theta_{j,d_j}]^\top$$

$$\theta = [\theta_{1,1} \cdots \theta_{1,d_1} \theta_{2,1} \cdots \theta_{2,d_2} \cdots \theta_{p,1} \cdots \theta_{p,d_p}]^\top \in \mathbb{R}^d$$

Linear regression on raw features

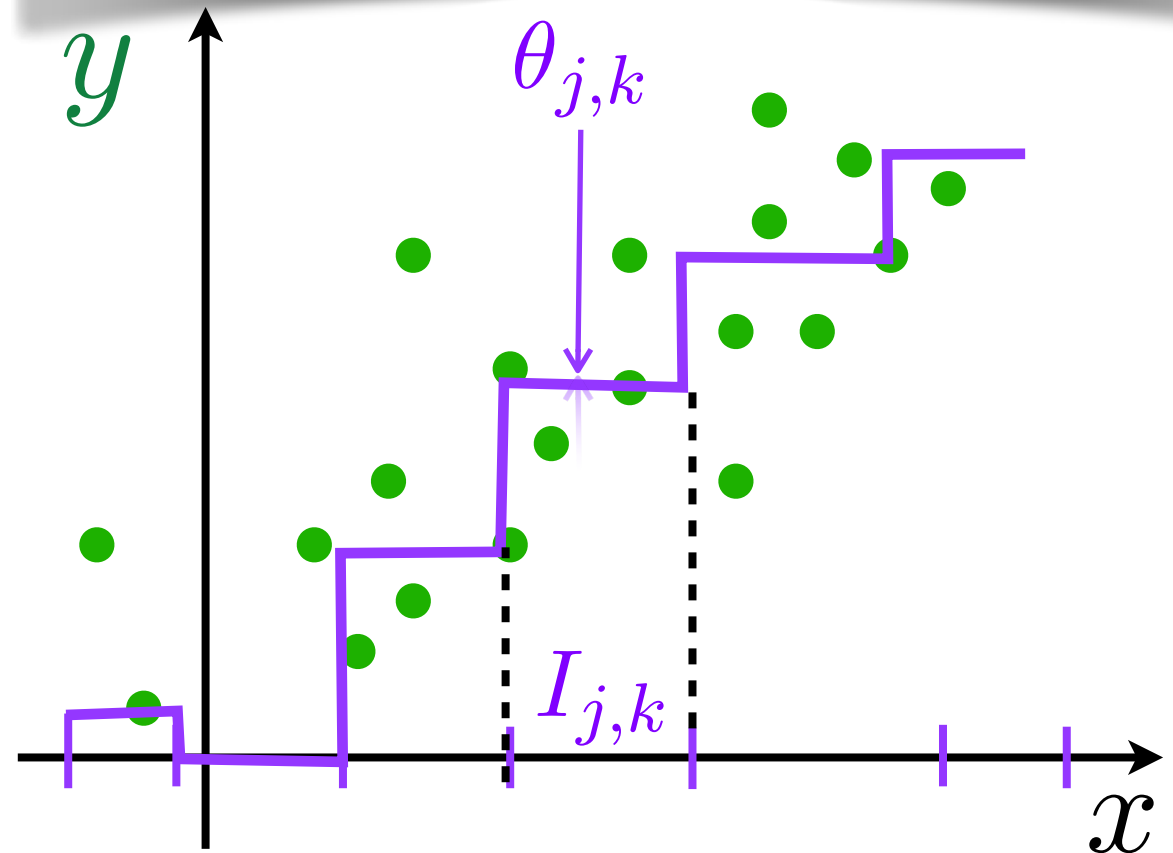


$$y_i = \omega^\top x_i + b = \sum_{j=1}^p \omega_j x_{i,j} + b$$

Impact of the j -th feature is **linear** and encoded by a **single weight** ω_j

$$x \mapsto \omega_j x$$

Linear regression on binarized features



$$y_i = \sum_{j=1}^p \sum_{k=1}^{d_j} \theta_{j,k} x_{i,j,k} + b = \sum_{j=1}^p \sum_{k=1}^{d_j} \theta_{j,k} \mathbb{1}(x_{i,j} \in I_{j,k}) + b$$

Impact of the j -th feature is **piecewise constant** and encoded by a **block**

$$\theta_{j,\bullet} = [\theta_{j,1} \cdots \theta_{j,d_j}]^\top$$

$$x \mapsto \sum_{k=1}^{d_j} \theta_{j,k} \mathbb{1}(x \in I_{j,k})$$

Pénalisation Binarsity

$$\text{bina}(\theta) = \sum_{j=1}^p \left(\sum_{k=2}^{d_j} \hat{\omega}_{j,k} |\theta_{j,k} - \theta_{j,k-1}| \right) + \delta_j(\theta_{j,\bullet})$$

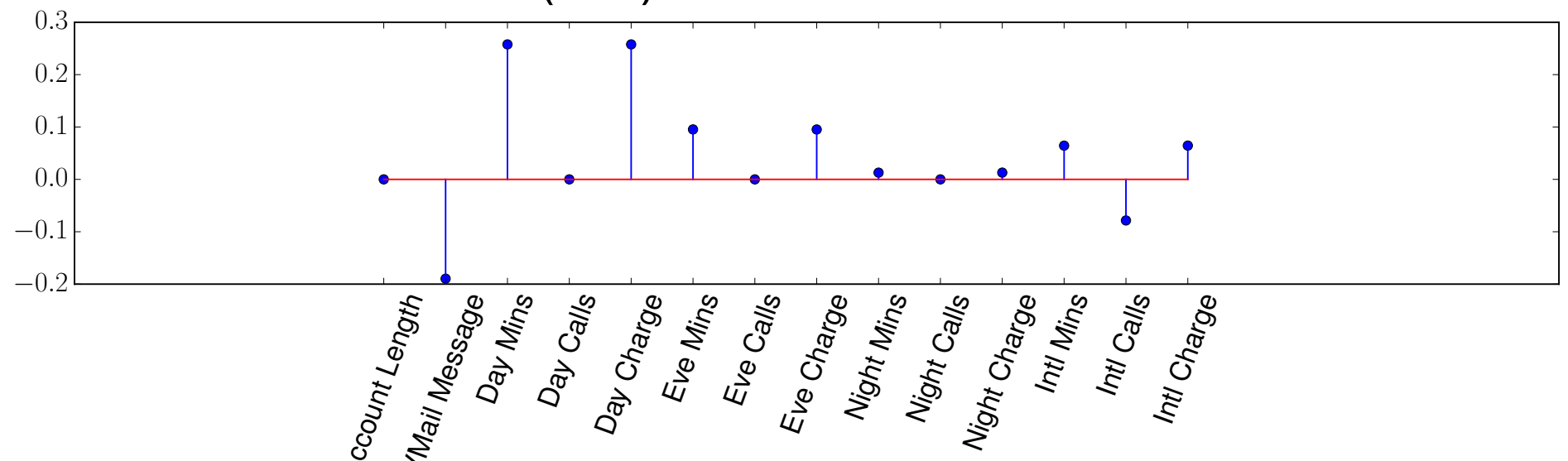
avec $\delta_j(\theta_{j,\bullet}) = \begin{cases} 0 & \text{si } n_j^\top \theta_{j,\bullet} = 0 \\ +\infty & \text{sinon} \end{cases}$ $n_j = [n_{j,1} \cdots n_{j,d_j}] \in \mathbb{N}^{d_j}$ $n_{j,k} = |\{i : X_{i,j} \in I_{j,k}\}|$

Poids

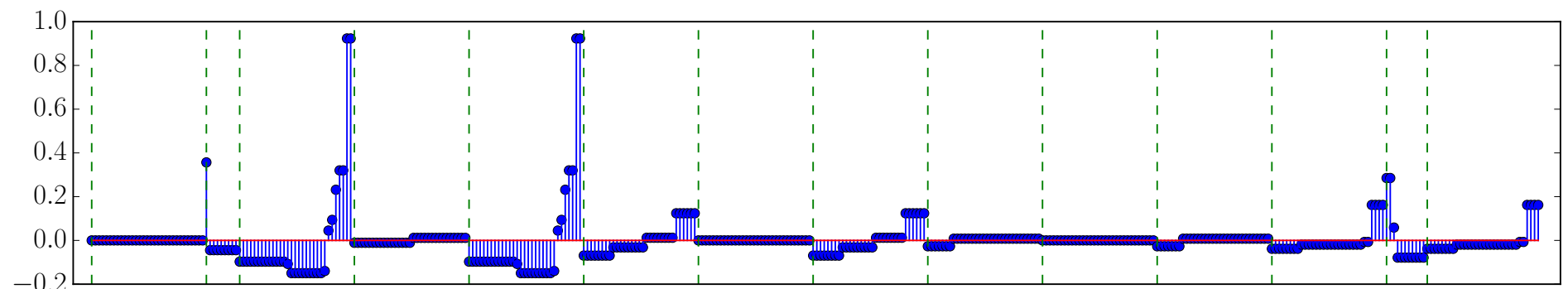
$$\hat{\omega}_{j,k} \approx \sqrt{\frac{\log d}{n}} \hat{\pi}_{j,k} \quad \hat{\pi}_{j,k} = \frac{|\{i=1, \dots, n : X_{i,j} \in \cup_{k'=k}^{d_j} I_{j,k'}\}|}{n}$$

Churn dataset (UCI) $n = 3333, p = 14$

Régression logistique
Features d'origine +
Lasso



Régression logistique
Features binaires +
pénalisation forte de
binarsity



Binarsity: Interprétation

Pénalisation

 ℓ_1

Sparsité

Pénalisation

*Binarsity*Constant par morceaux $\theta_{j,\bullet}$ Si $\theta_{j,\bullet}$ est constant par morceaux alors $\theta_{j,\bullet} \equiv 0$

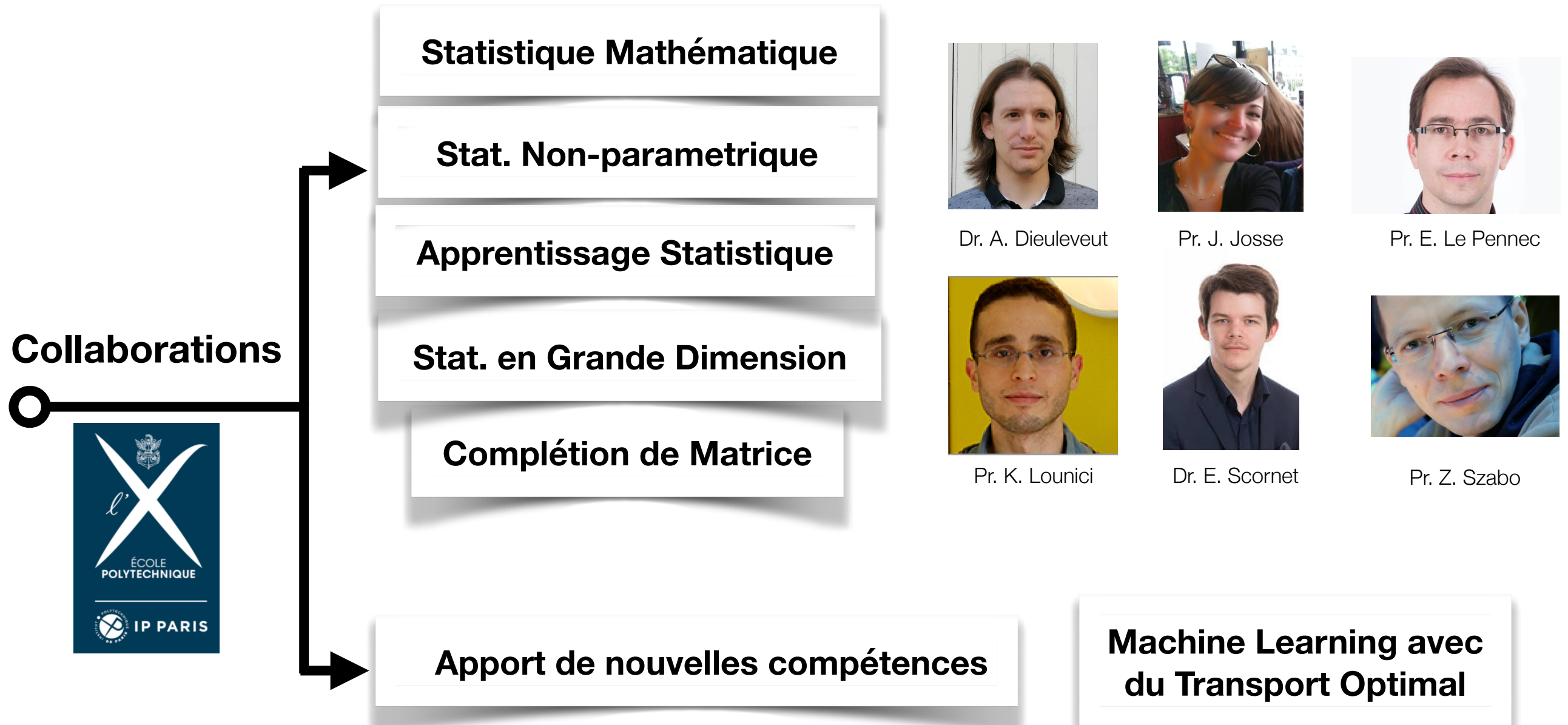
Bloc-Sparsité

 $\theta_{j,\bullet} \neq 0$ $\theta_{j,\bullet} \equiv 0$ $X_{\bullet,j}$

Pertinente pour la prédiction si

Non-Pertinente pour la prédiction si

Intégration dans l'Équipe SIMPAS



Rouen, le 1 juin 2020



Merci pour votre attention!