Méthodes d'Apprentissage Automatique pour l'Aide au Diagnostic

Yohan Dupuis

Bienvenue

Instructeur

- **Yohan Dupuis, PhD, MScEE, Meng**,
  - Chef de Projet, Chercheur MEDDE
  - CEREMA DTerNC
  - E-mail : *yohan.dupuis@cerema.fr,*
  - Tél : 02 35 68 90 55,
  - Contact : E-mail ou sur RDV.

Plan de cours

I. Introduction
  I.Notions de base
  II.Statistiques – Probabilités – Algèbre Linéaire

II. Apprentissage Supervisé
  I.Régression
  II.Classification

III. Apprentissage Non-Supervisé

Notation

Quizz : 20 %

Projet : 50 %

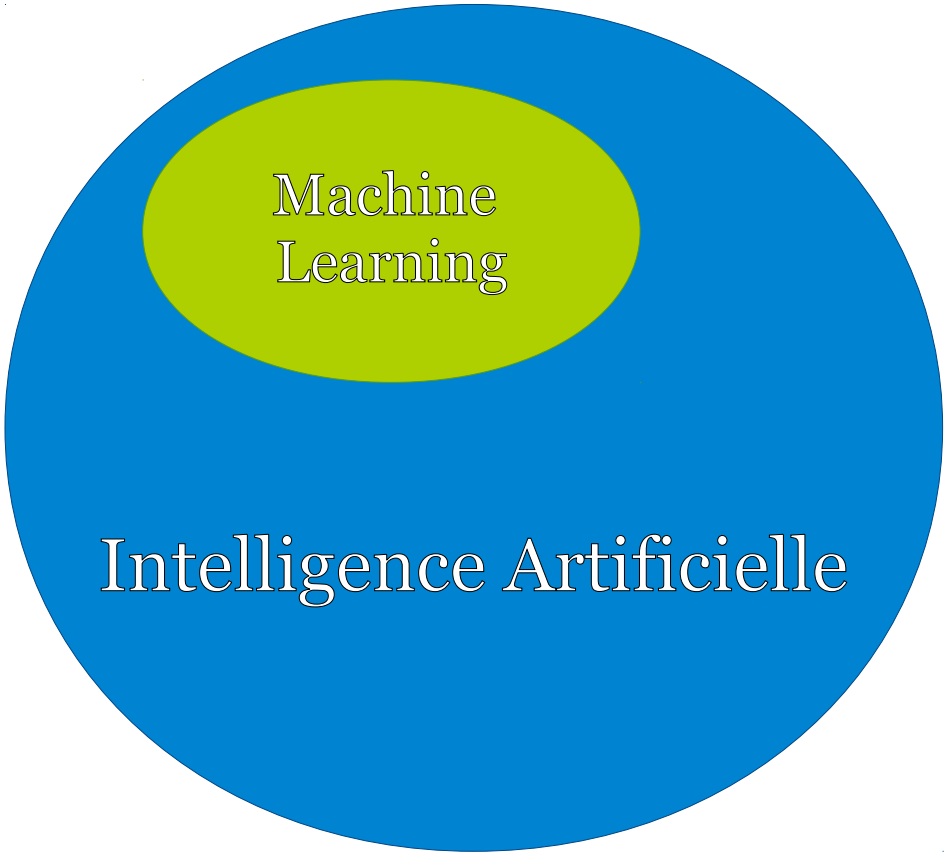
Examen: 30 %

Sources

- Statistics, *San Jose State University, USA*.
- Linear Algebra, *MIT, USA*.
- Data Mining, *Union Graduate College, USA*.
- Applied Machine Learning, *Stanford, USA*.

Introduction

Machine Learning



- Data Mining
- Application complexes
- Conseils personnalisés
- Compréhension de l'apprentissage

Définitions

« Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed »

- Arthur Samuel (1901-1990)

Définitions

*« A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E** »*

-Tom Michael Mitchel (1961-)

Familles d'algorithmes

- Apprentissage supervisé
- Apprentissage non supervisé
- Apprentissage par renforcement
- Systèmes de recommandation

Introduction

Apprentissage Supervisé

Coûts & Santé

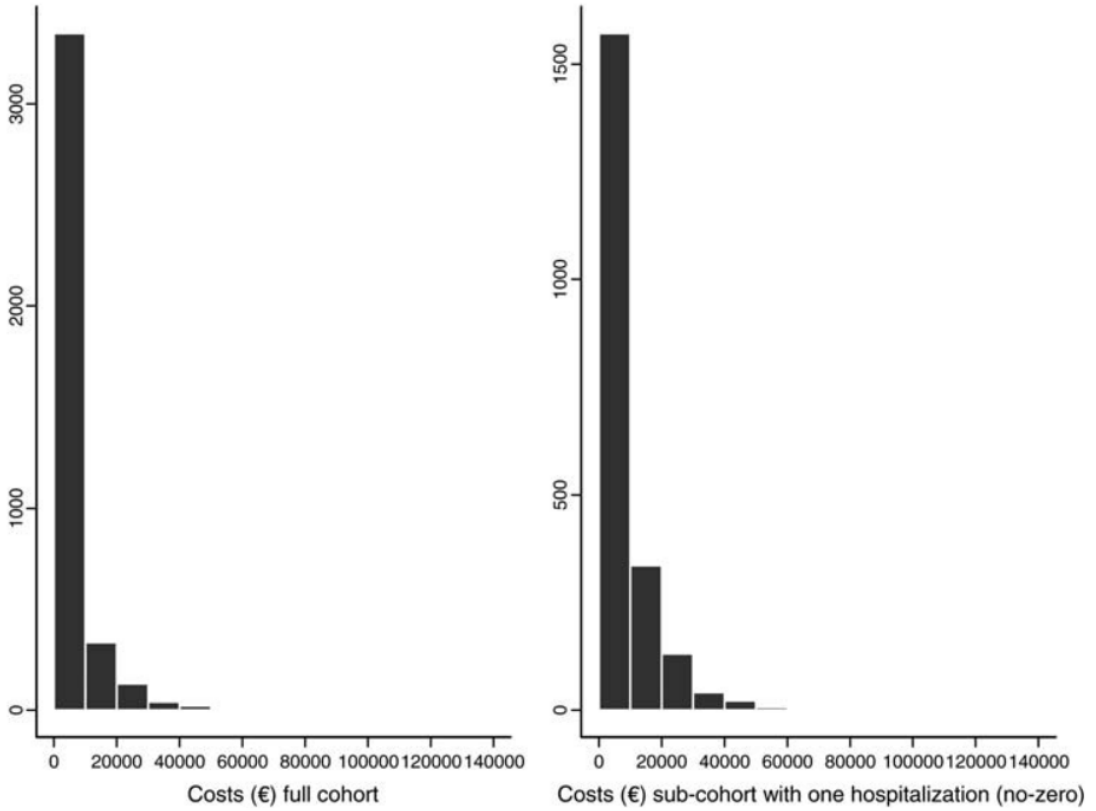


Figure 1 Cost distribution for the diabetes cohort, with zero-cost patients (left plot) and without (right plot).

Source : «Regression models for analyzing costs and their determinants in health care: an introductory review», International Journal for Quality in Health Care, 2011

Coûts & Santé

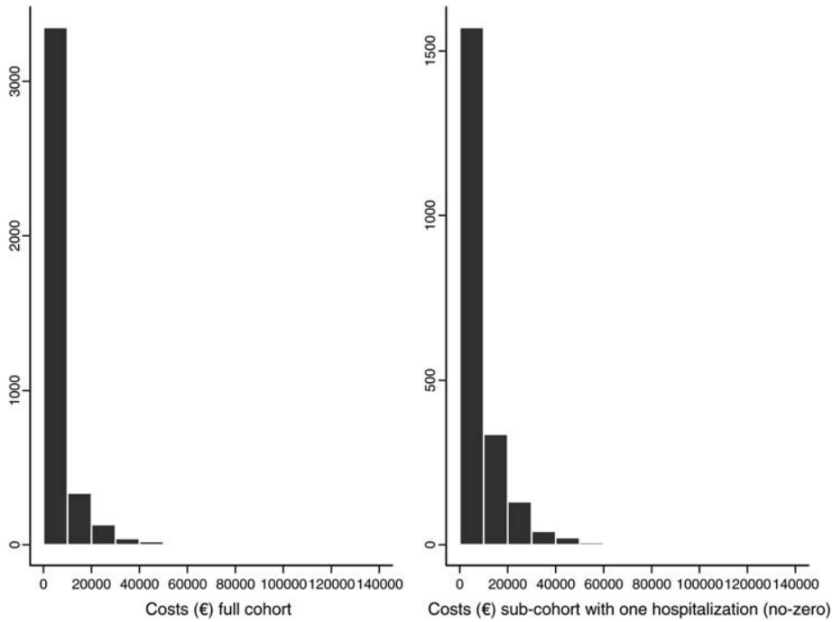


Figure 1 Cost distribution for the diabetes cohort, with zero-cost patients (left plot) and without (right plot).

- Supervisé : connaissance de valeurs « vraies » .
- Régression : prédiction d'une valeur continue.

Source : «Regression models for analyzing costs and their determinants in health care: an introductory review», International Journal for Quality in Health Care, 2011

Reconnaissance d'activités

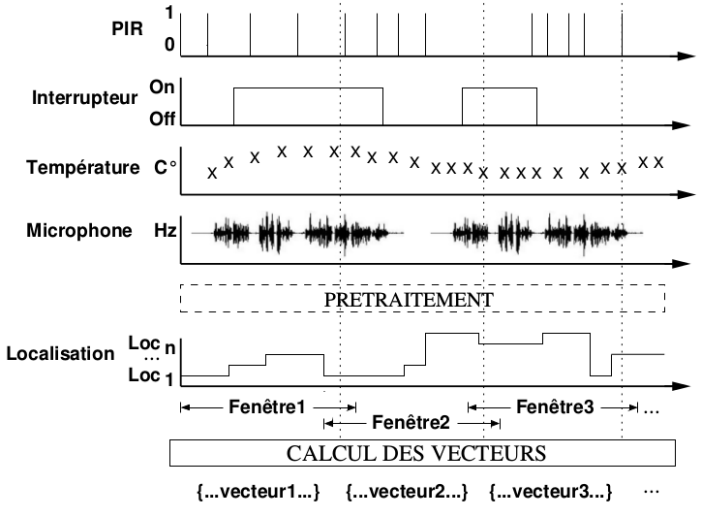


FIGURE 1 – Fenêtrage temporel utilisé pour l'obtention des attributs à partir des données fournies par les capteurs.

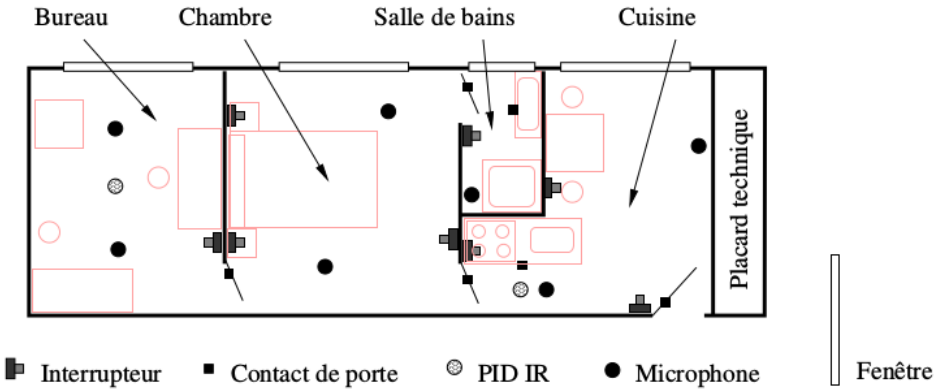


FIGURE 2 – Plan de l'appartement DOMUS et disposition des capteurs.

Source : «Méthodes SVM et MLN pour la reconnaissance automatique d'activités humaines dans les habitats perceptifs: tests et perspectives», RFIA , 2012

Reconnaissance d'activités

Rejet	Sans						Avec					
Fenêtrage	60 secs			120 secs			60 secs			120 secs		
Recouvrement	0	25%	50%	0	25%	50%	0	25%	50%	0	25%	50%
Repas	75,4	74,3	72,9	75,9	75,3	73,6	73,3	74,3	75	74,1	71,8	75,3
Ranger	50,3	51,9	52,1	56,8	55,6	55	42,2	46,9	49,5	49,2	50	49,6
Hygiène	44	40,7	35,5	40,9	40,9	40,9	32	25,9	25,8	40,9	40,9	40,9
Détente	79,9	80,6	82	81,6	80,1	82,1	76,3	78	81,4	80	79,4	80,8
Habillage	4,8	4,8	4,8	4,8	2,4	9,5	2,4	0	4,8	4,8	4,8	2,4
Dormir	68,1	81,8	83,3	63,8	60,8	63,2	68,1	78,4	82,5	55,3	62,7	61,4
Rejet	–	–	–	–	–	–	9,3	7,9	7,3	17,1	9,8	7,3

TABLE 3 – Rappel de la classification des activités avec les SVM (validation croisée à 5 partitions)

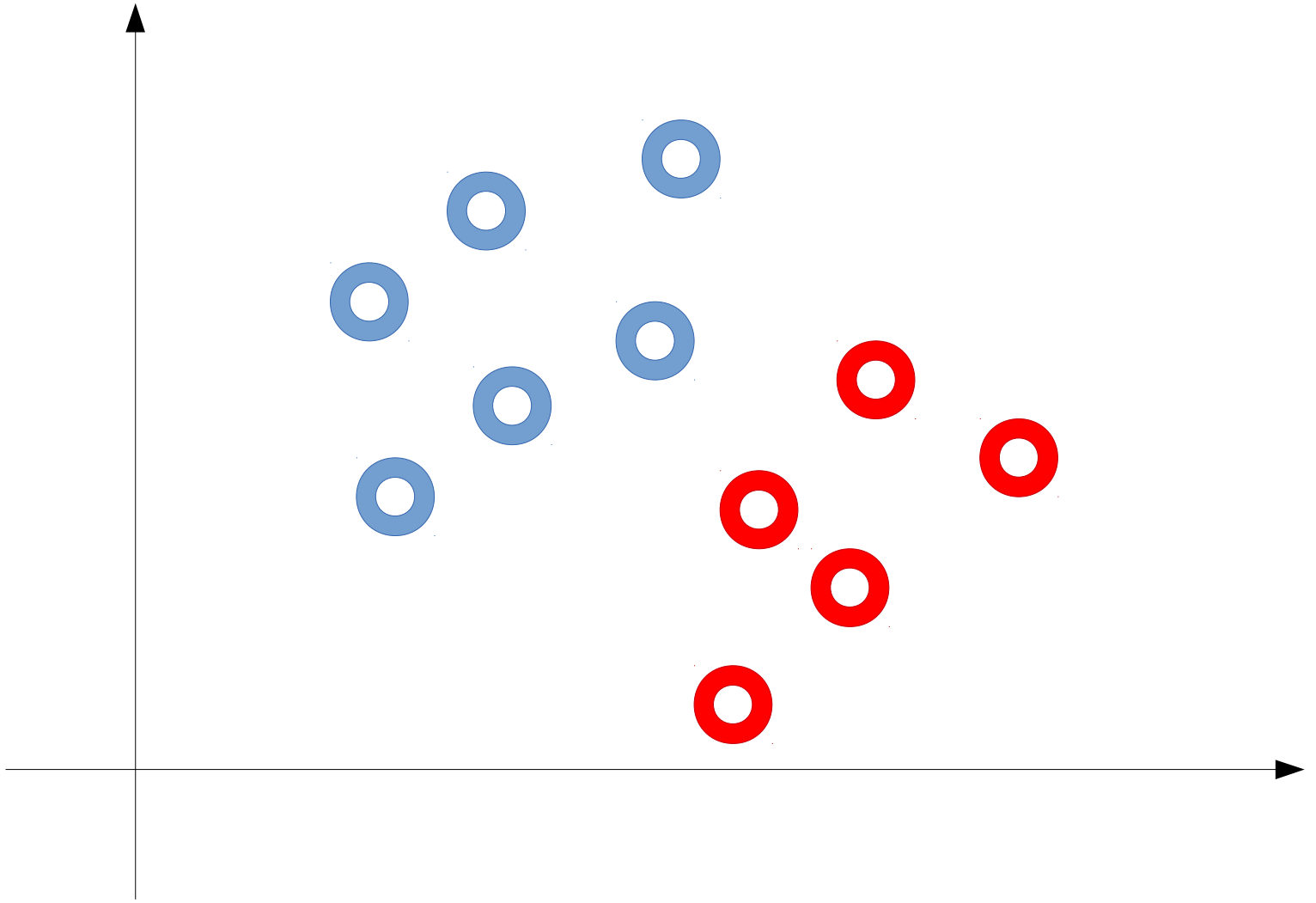
- Supervisé : connaissance de valeurs « vraies ».
- Classification : prédiction d'une valeur discrète.

Source : «Méthodes SVM et MLN pour la reconnaissance automatique d'activités humaines dans les habitats perceptifs: tests et perspectives», RFIA , 2012

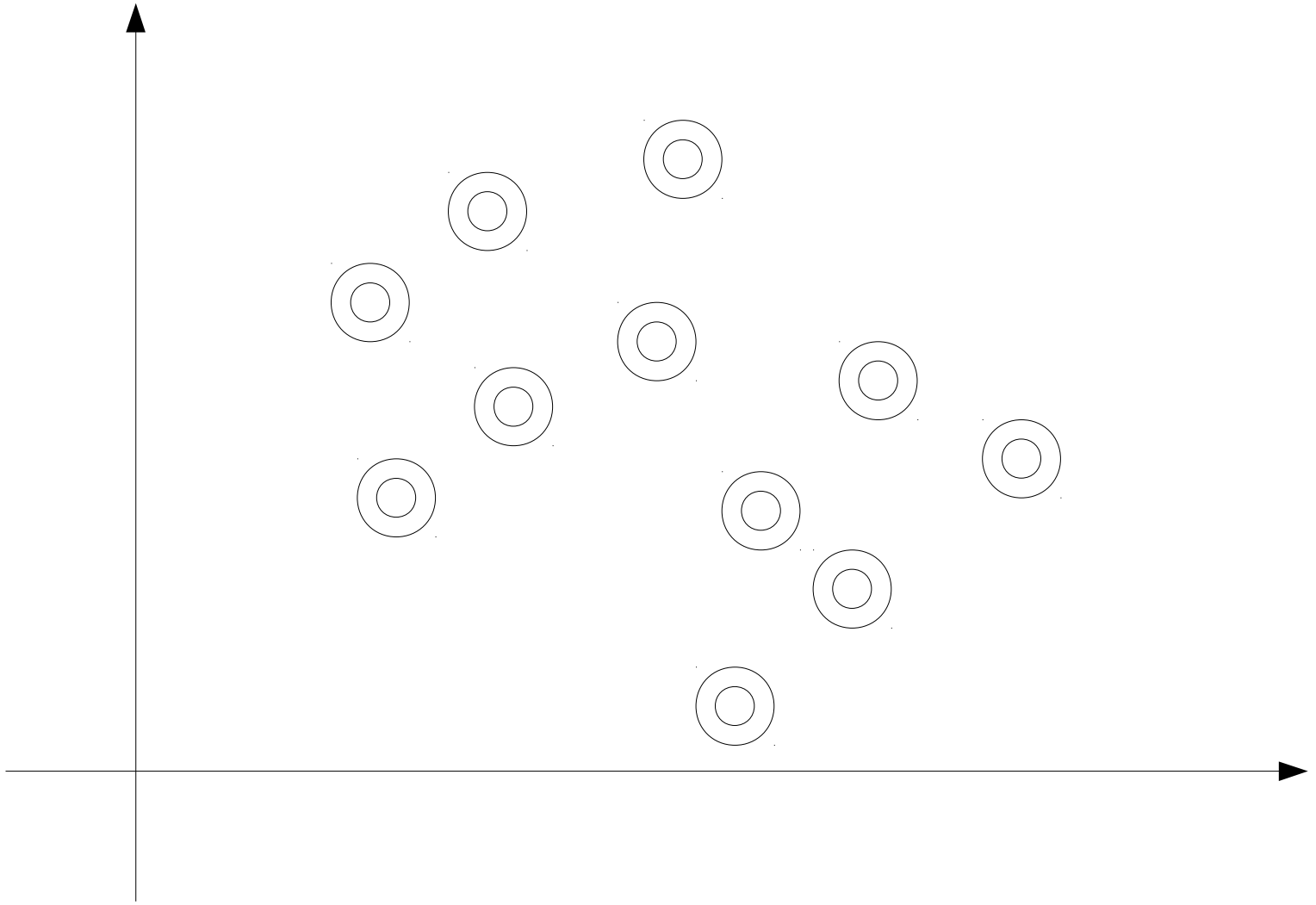
Introduction

Apprentissage Non Supervisé

Apprentissage supervisé



Apprentissage non supervisé



Objectif ?

Marketing - Recommandations

Marketing - Recommandations

Saint-Valentin
Des livres à offrir ou à partager
›Découvrir

Voir les 2 images

Ajouter au panier

Identifiez-vous pour activer la commande 1-Click.

Ajoutez à votre liste d'envies

Amazon rachète votre article **EUR 12,50** en chèque-cadeau. Revendre

En savoir plus

Vous l'avez déjà ? Vendez sur Amazon

Partager

Offres spéciales et liens associés

- -40%, -50%, -60%, -70%... Découvrez les **Soldes Amazon** jusqu'au 17 février 2015 inclus. **Profitez-en !**

Produits fréquemment achetés ensemble

Prix pour les trois: EUR 46,80

Ajouter ces trois articles au panier

Afficher la disponibilité du produit et le mode de livraison

- **Cet article :** Le capital au XXIe siècle de Thomas Piketty Broché EUR 25,00
- L'économie des inégalités de Thomas PIKETTY Broché EUR 10,00
- Pour une révolution fiscale : Un impôt sur le revenu pour le XXIe siècle de Camille Landais Broché EUR 11,80

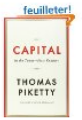
Les clients ayant acheté cet article ont également acheté

Page 1 sur 1

L'économie des inégalités
Thomas PIKETTY
Broché
EUR 10,00

Le triomphe de la cupidité
Joseph Stiglitz
(23)
Broché
EUR 10,70

Soumission
› Michel Houellebecq
(163)
Broché
EUR 21,00



Capital in the Twenty-First Century
› Thomas Piketty
(5)
Relié
EUR 32,65

Pour une révolution fiscale : Un impôt sur le revenu ...
Camille Landais
(23)
Broché
EUR 11,80

L'économie des inégalités
Thomas PIKETTY
(9)
Broché

Les hauts revenus en France au XXème siècle Ned
› Thomas Piketty
Broché
EUR 31,00

La Grande Désillusion
› Joseph E. Stiglitz
(26)
Poche
EUR 6,90

Introduction à la théorie de la redistribution des richesses
› Thomas Piketty
Broché
EUR 15,00

Descriptions du produit

Bio-informatique

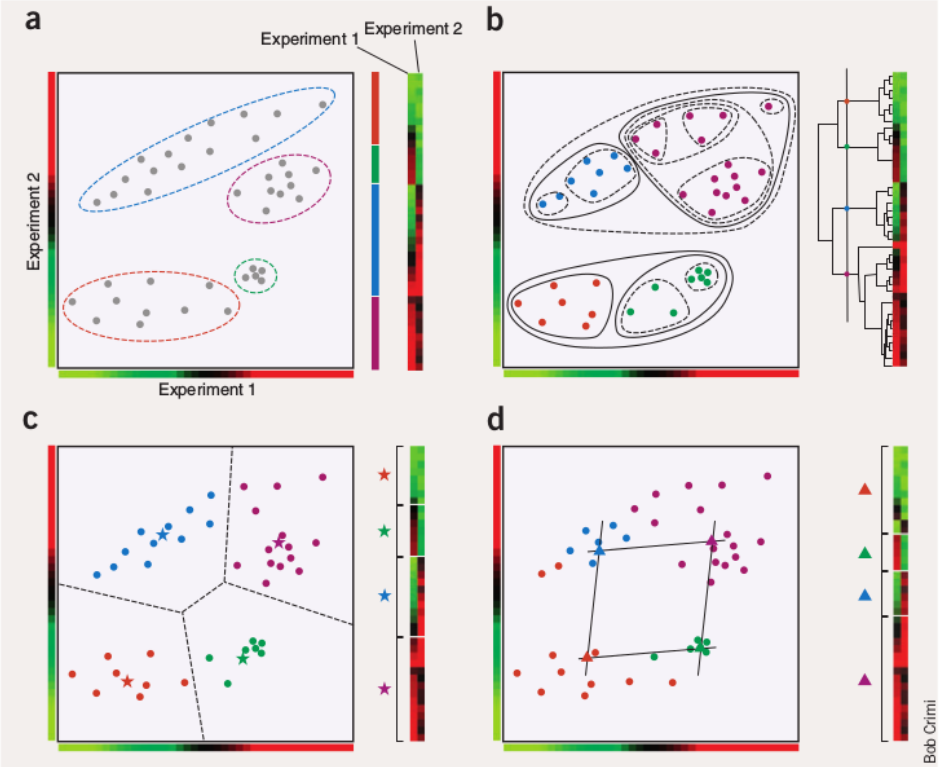


Figure 1 A simple clustering example with 40 genes measured under two different conditions. (**a**) The data set contains four clusters of different sizes, shapes and numbers of genes. Left: each dot represents a gene, plotted against its expression value under the two experimental conditions. Euclidean distance, which corresponds to the straight-line distance between points in this graph, was used for clustering. Right: the standard red-green representation of the data and corresponding cluster identities. (**b**) Hierarchical clustering finds an entire hierarchy of clusters. The tree was cut at the level indicated to yield four clusters. Some of the superclusters and subclusters are illustrated on the left. (**c**) *k*-means (with $k = 4$) partitions the space into four subspaces, depending on which of the four cluster centroids (stars) is closest. (**d**) SOM finds clusters, which are organized into a grid structure (in this case a simple 2 × 2 grid).

Source : «How does gene expression clustering work?», NATURE BIOTECHNOLOGY , 2005

Introduction

Algèbre Linéaire

DATA

Organisation des données

b=5

d=-50

a=10

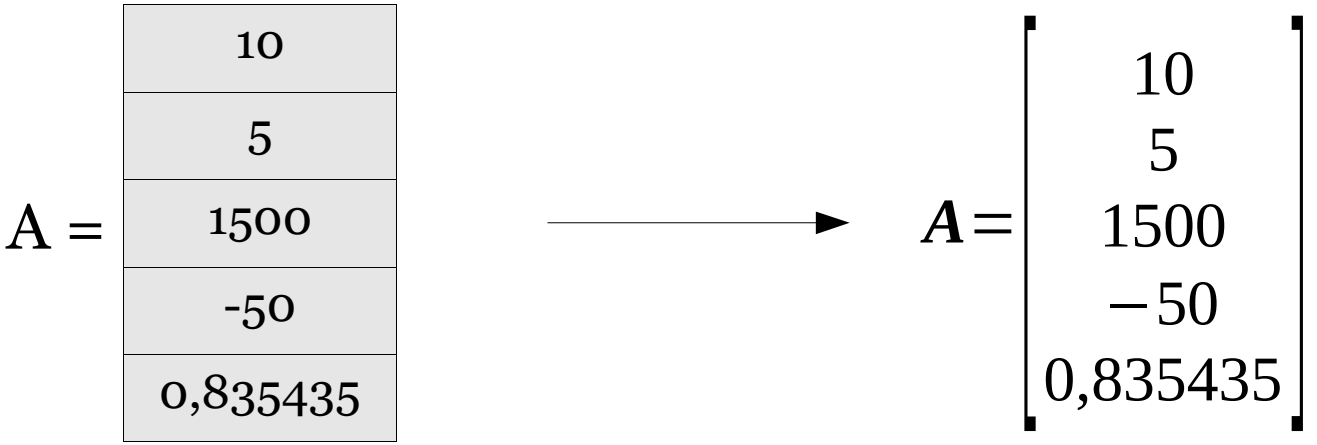
e=0,835435

c=1500

A =

10
5
1500
-50
0,835435

Organisation des données



Notations

- Scalaire = nombre

$$a = 10$$

- Vecteur ligne

$$\boldsymbol{b} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

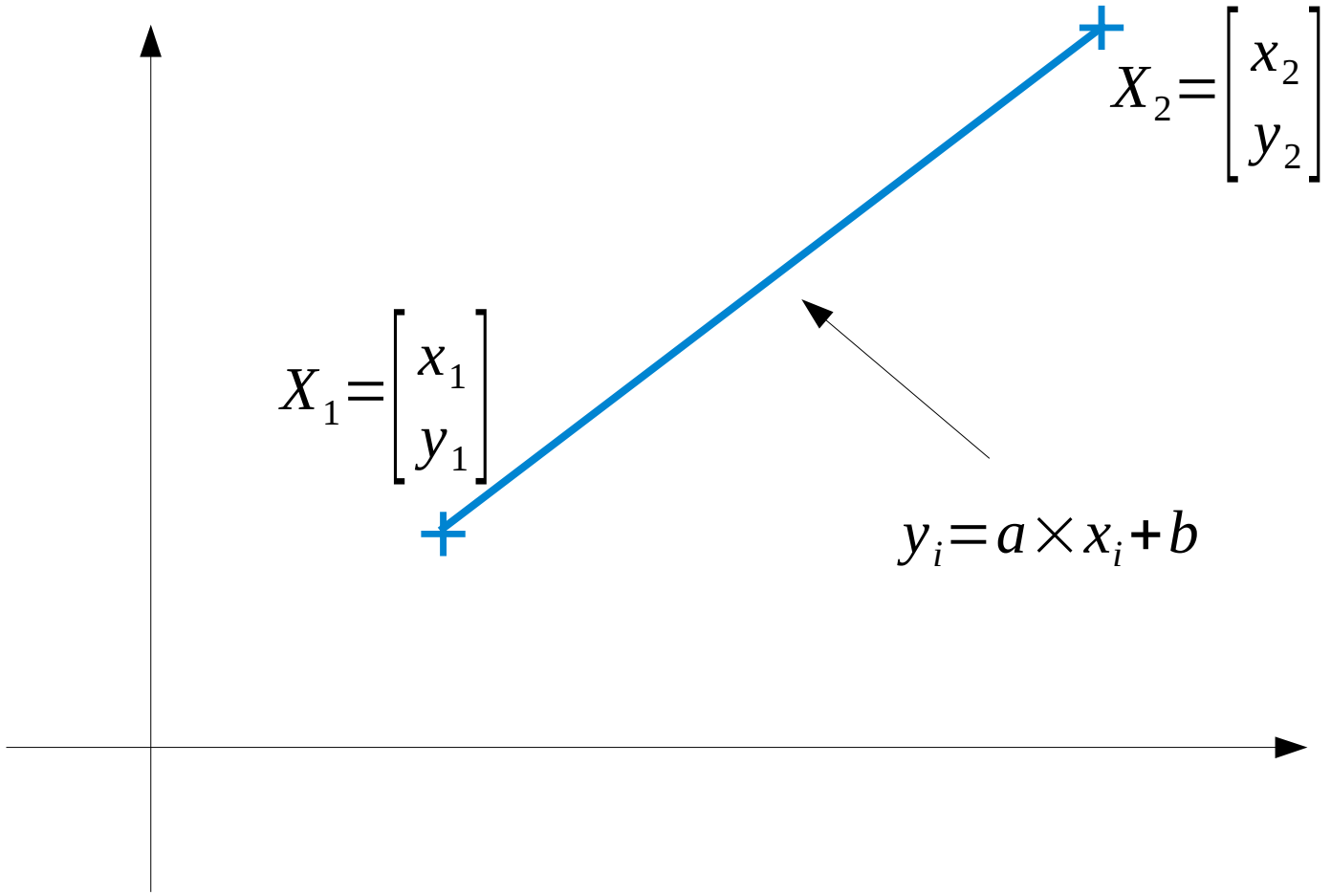
- Vecteur colonne

$$\boldsymbol{c} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \boldsymbol{c} = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$$

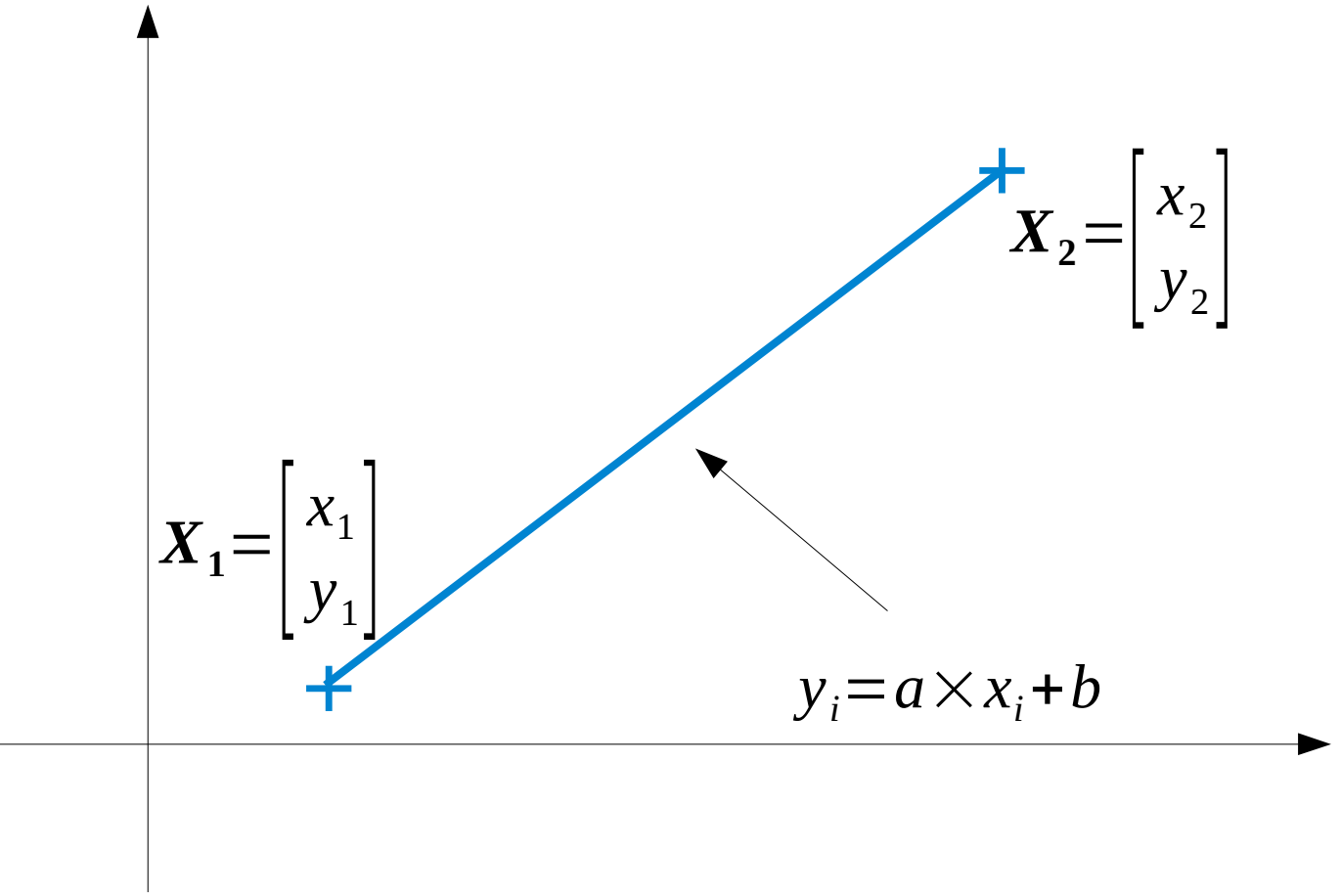
- Matrice

$$\boldsymbol{D} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Exemple


$X_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$
$X_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$
$y_i = a \times x_i + b$

Exemple



$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

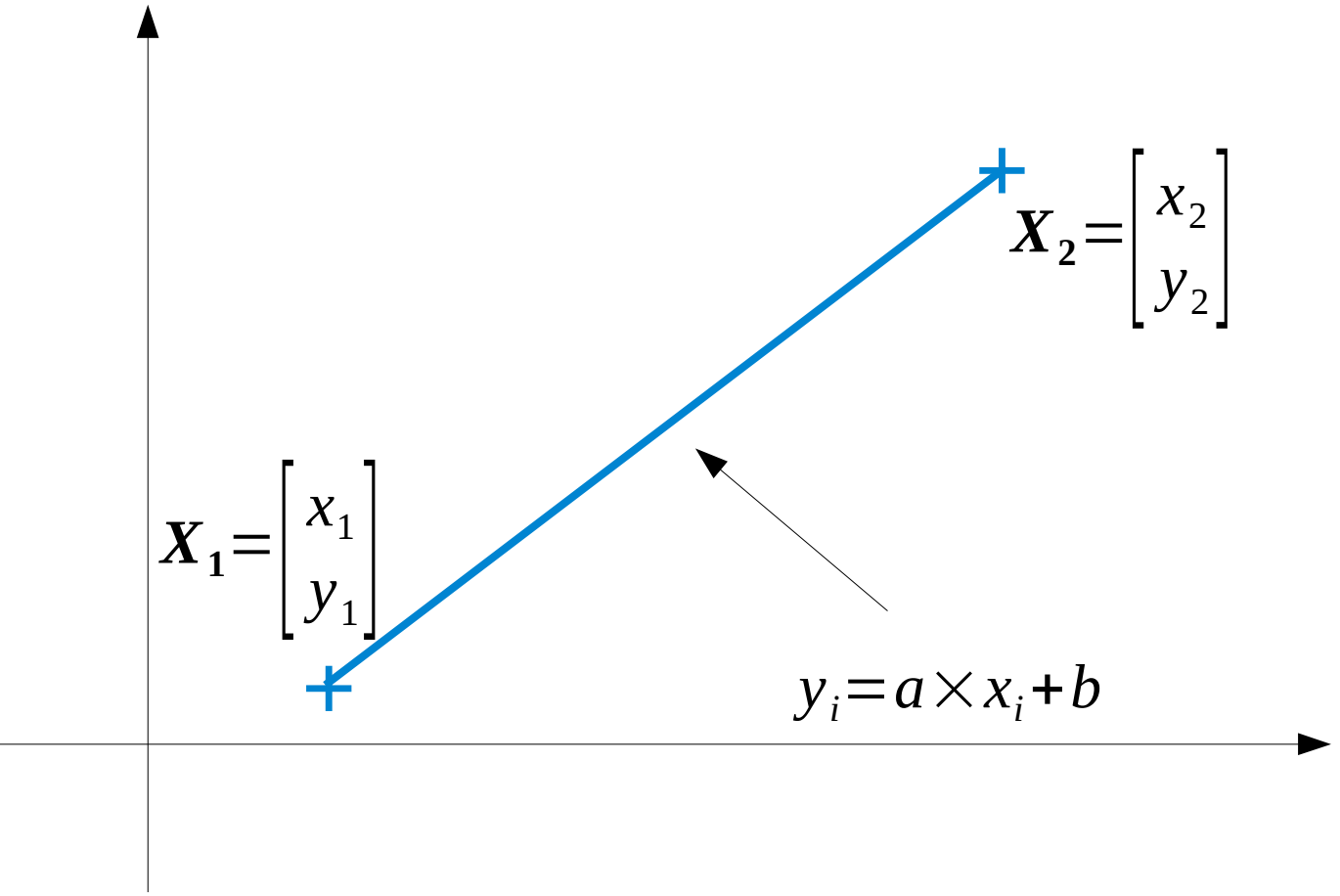
$$\boldsymbol{Ax} = \boldsymbol{b}$$

$$\boldsymbol{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} a \\ b \end{bmatrix}$$

Exemple



$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\boldsymbol{Ax = b}$$

$$\boldsymbol{x = A^{-1} \times b}$$

Merci de votre Attention

Questions

Démo Matlab