

# Apprentissage statistique

## Généralités

Vincent Lefieux

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références



# Plan

Introduction

Introduction

Apprentissage supervisé

Apprentissage  
supervisé

Sur-apprentissage

Sur-  
apprentissage

Critères d'évaluation de la performance de modèles

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur la classification supervisée

Compléments sur  
la classification  
supervisée

Références

# Plan

## Introduction

Introduction

## Apprentissage supervisé

Apprentissage  
supervisé

## Sur-apprentissage

Sur-  
apprentissage

## Critères d'évaluation de la performance de modèles

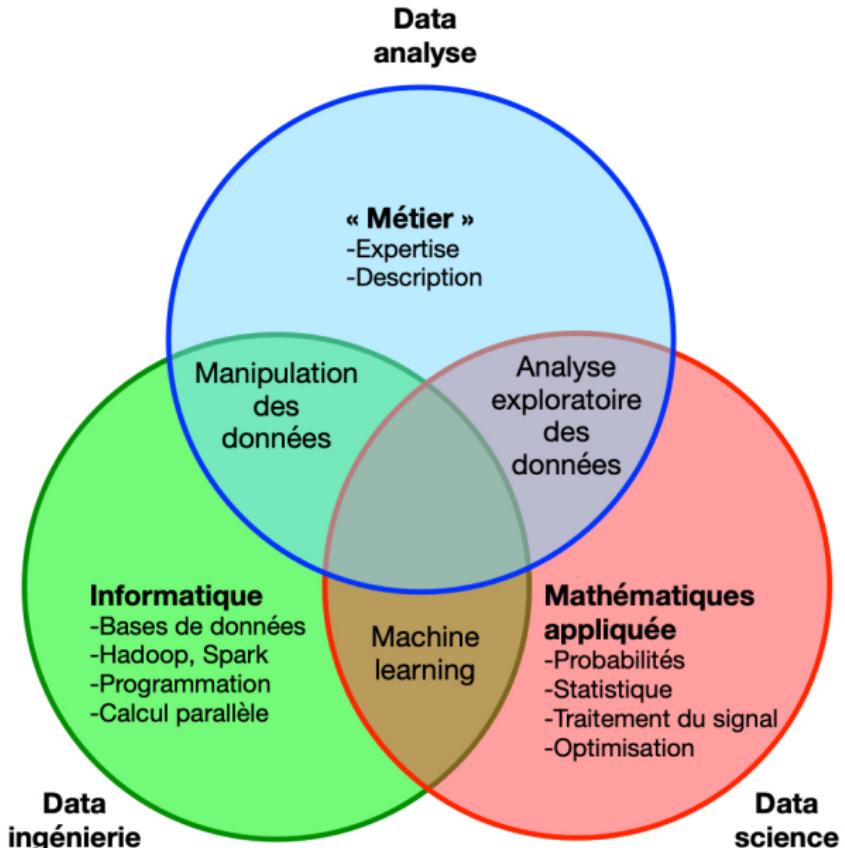
Critères  
d'évaluation de  
la performance  
de modèles

## Compléments sur la classification supervisée

Compléments sur  
la classification  
supervisée

Références

# De la data science



Introduction

Apprentissage supervisé

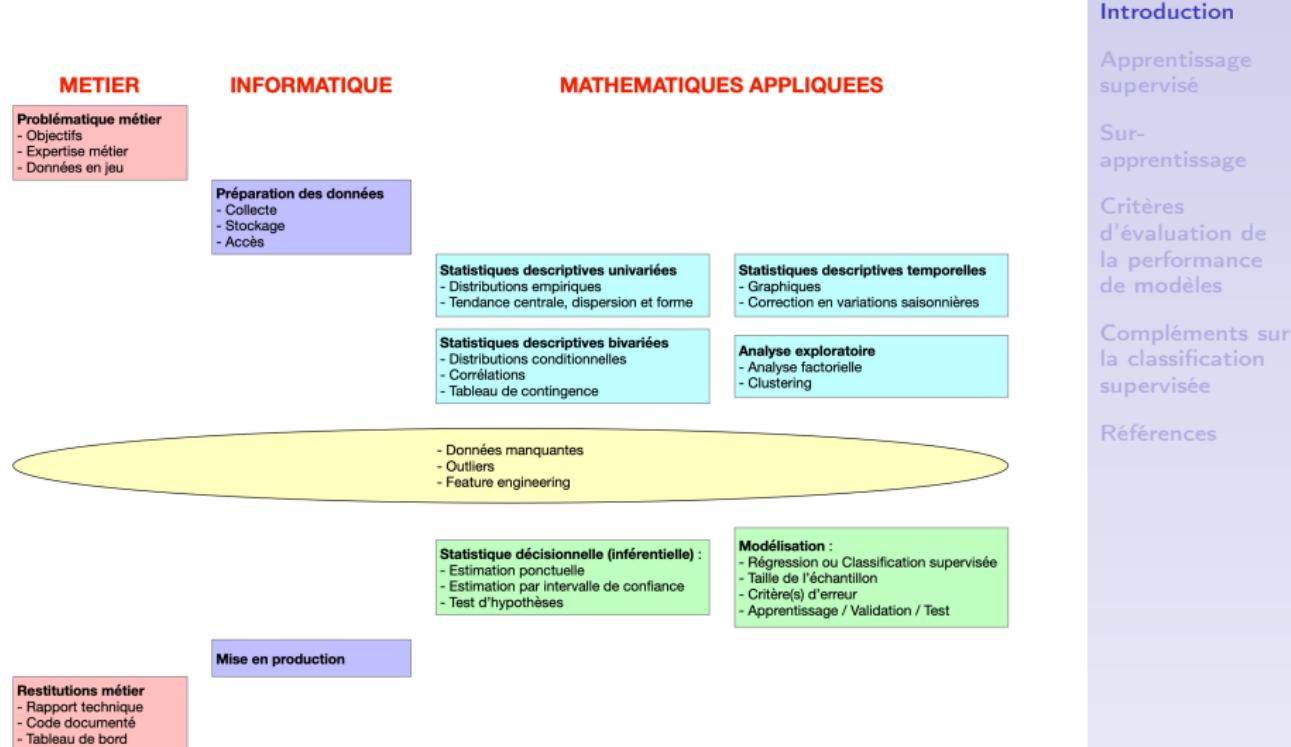
Sur-apprentissage

Critères d'évaluation de la performance de modèles

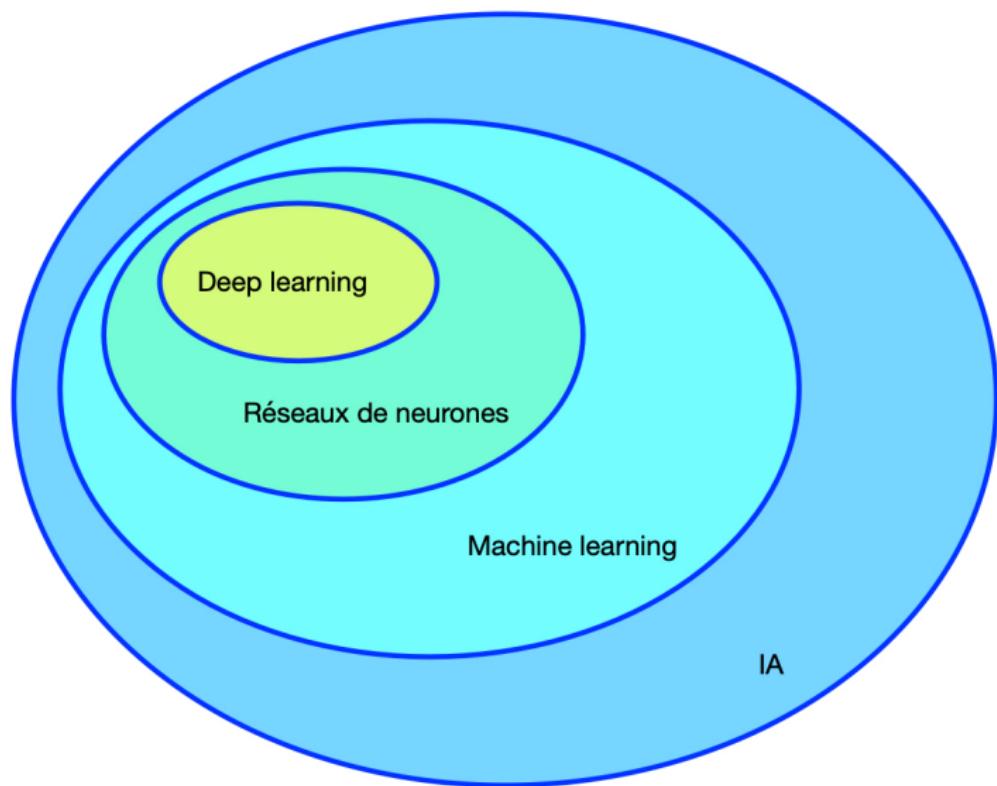
Compléments sur la classification supervisée

Références

# Parcours « data »



# De l'IA



Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Différents types d'apprentissage I

## ► Apprentissage supervisé :

Inférer (prédire) une fonction ou une relation à partir de données d'apprentissage labellisées (on parle aussi d'exemples étiquetés).

On distingue :

- ▶ La régression : pour un label quantitatif.
- ▶ La classification supervisée : pour un label qualitatif.

## ► Apprentissage non-supervisé :

Trouver une « structure » dans des données non-labellisées (ex : clustering, *dimension reduction*).

Même s'il est plus « subjectif » que l'apprentissage supervisé, il peut être utile comme étape de pré-traitement pour l'apprentissage supervisé.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Différents types d'apprentissage II

- ▶ Apprentissage **par renforcement** :  
Faire apprendre à un agent autonome les actions à prendre à partir d'expériences, de manière à optimiser une récompense quantitative au cours du temps.
- ▶ Apprentissage **par transfert** :  
Transposer par analogie un apprentissage mené sur un problème similaire mais différent.

Introduction

Apprentissage supervisé

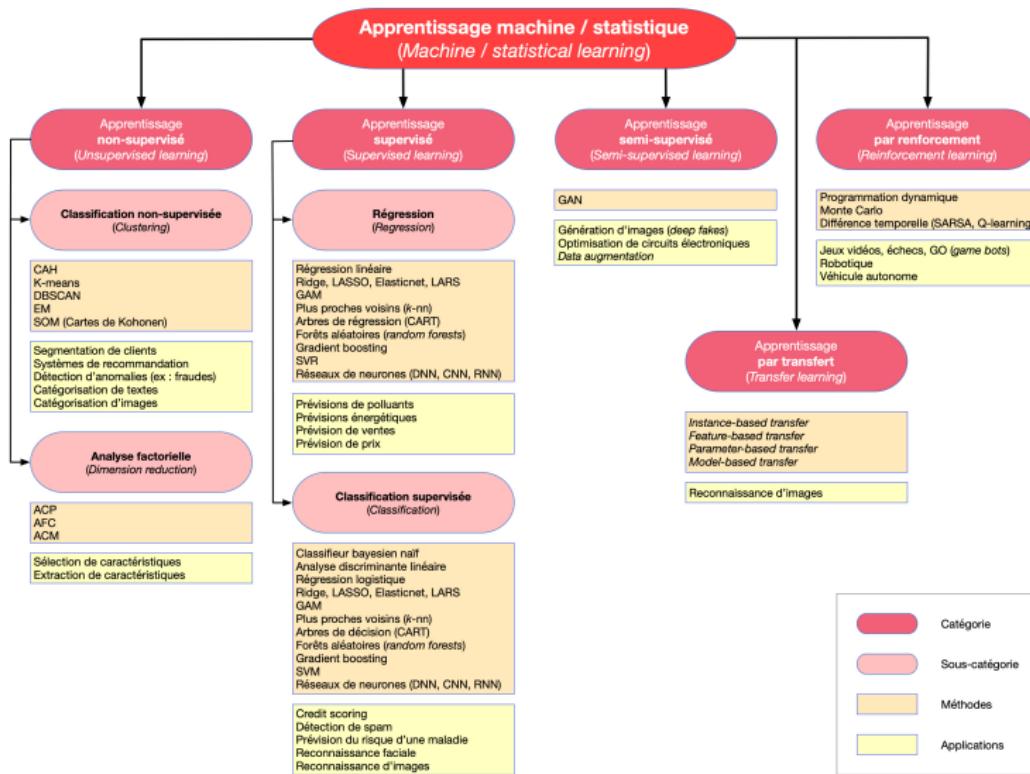
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Différents types d'apprentissage III



Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

- Catégorie
- Sous-catégorie
- Méthodes
- Applications

# Quelques méthodes d'apprentissage supervisé

- ▶ **Modèle linéaire généralisé** : régression linéaire, régression de Poisson, régression logistique.
- ▶ **Méthodes régularisés** : Ridge, Lasso, Elasticnet, Lars.
- ▶ **Méthodes bayésiennes**.
- ▶ **Méthodes splines** : régression spline, GAM.
- ▶ D'autres méthodes : analyse discriminante linéaire, PLS, méthodes à directions révélatrice (*index model*).
- ▶ **Méthodes de moyennage local** : plus proches voisins, noyau de lissage, CART.
- ▶ **Méthodes d'agrégation** : bagging (dont random forests), gradient boosting.
- ▶ **Méthodes à noyau** : SVM (et SVR).
- ▶ **Réseaux de neurones** : DNN, CNN, RNN.

# Quelques méthodes d'apprentissage non-supervisé

- ▶ **Classification non supervisé** (*clustering*) : K-means, CAH, DBSCAN, modèles de mélange, auto-encodeurs, cartes de Kohonen (SOM : Self Organizing Maps).
- ▶ **Réduction de dimension** (*dimension reduction*) : ACP, AFC, ACM.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Modélisation et/ou prévision

- ▶ On peut distinguer **modélisation** et **prévision**, par exemple compression d'image vs reconnaissance d'images.
- ▶ Un modèle s'appuie sur la **régularité** des phénomènes sous-jacents.
- ▶ La **prévision** consiste à **généraliser** un modèle.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Quelques enjeux en prévision

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

- ▶ Compromis entre la **qualité de la prévision** et l'**interprétabilité** (notion de « boîte noire »).
- ▶ Privilégier des **modèles parcimonieux** (« *sparse* ») qui éviteront le **sur-apprentissage** : *less is more*.

# Plan

Introduction

## Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

## Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plan

Introduction

**Apprentissage supervisé**

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

**Formalisation**

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Données

Introduction

Apprentissage  
supervisé

**Formalisation**

Pertes et risques

Biais et variance

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

- ▶ On dispose d'un échantillon de  $(X, Y)$  :

$$\mathcal{D}_n = (X_i, Y_i)_{i \in \{1, \dots, n\}}$$

où  $X \in \mathcal{X}$  et  $Y \in \mathcal{Y}$ .

- ▶ On note :

$$d_n = (x_i, y_i)_{i \in \{1, \dots, n\}} .$$

# Objectif

On se placera dans le cadre de la **prévision** : on souhaite prévoir  $y$  pour une nouvelle valeur  $x$ .

Introduction

Apprentissage  
supervisé

**Formalisation**

Pertes et risques

Biais et variance

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

# Covariables

Introduction

Apprentissage  
supervisé

**Formalisation**

Pertes et risques

Biais et variance

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

- ▶ On considèrera très souvent dans la suite que :

$$X \in \mathbb{R}^p .$$

- ▶ Par défaut les covariables seront considérées comme quantitatives mais on pourra traiter des variables qualitatives :
  - ▶ directement dans certaines méthodes (ex : arbres),
  - ▶ à l'aide d'indicatrices sur chacune des modalités des variables qualitatives).

# Régression et classification supervisée

Introduction

Apprentissage  
supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

- ▶ **Régression** : la variable  $Y$  est quantitative.

Dans la suite on considérera que  $Y \in \mathbb{R}$ .

Mais il est possible de considérer plus généralement  
 $Y \in \mathbb{R}^d$ .

- ▶ **Classification supervisée** : la variable  $Y$  est qualitative.

Dans la suite on considérera que  $Y \in \{-1, 1\}$ .

Par défaut la classification supervisée sera considérée binaire mais on sera amené à traiter des classifications supervisées avec (strictement) plus de 2 labels.

# Prévision

- ▶ On suppose que  $(x_i, y_i)$  est la **réalisation** d'une v.a.r  $(X_i, Y_i)$  de loi de probabilité inconnue  $P_{X,Y}$  (modèle statistique non-paramétrique).
- ▶ La fonction de prévision de  $Y$  est une fonction  $f : \mathcal{X} \rightarrow \mathcal{Y}$ .
- ▶ On suppose que  $f \in \mathcal{F}$ .
- ▶ Dans la suite, de manière plus spécifique que  $f$ , on désignera la fonction de lien par :
  - ▶ Cas de la **classification supervisée** :  $g$  .
  - ▶ Cas de la **régression** :  $m$  .
- ▶ On cherche à **estimer**  $f$  par  $\hat{f}$ .

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plan

Introduction

**Apprentissage supervisé**

Formalisation

**Pertes et risques**

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Formalisation

**Pertes et risques**

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Qualité d'un prédicteur

- ▶ La qualité d'un prédicteur  $\hat{f}$  est évaluée par le **risque**  $R$  (ou encore **erreur de généralisation**) qui :
  - ▶ permet de sélectionner un modèle,
  - ▶ fournit un indice de la confiance qu'on peut avoir en une prévision.
- ▶ Le risque est défini à partir d'une **fonction de coût** (ou encore **fonction de perte**).

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Fonctions de perte

- ▶ On appelle **fonction de perte** une fonction  $\ell : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}^+$  telle que :

- ▶  $\ell(y, y) = 0$ ,
- ▶  $\forall y \neq y' : \ell(y, y') > 0$ .

- ▶ Exemples de fonctions de perte :

- ▶ Cas de la **classification supervisée binaire** :

$$\ell(y, y') = \mathbb{1}_{y \neq y'} = \frac{|y - y'|}{2} = \frac{(y - y')^2}{4}.$$

- ▶ Cas de la **régression** :

$$\ell(y, y') = |y - y'|^q$$

avec  $q \in \mathbb{R}^+$ .

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Risque (erreur de généralisation)

Le **risque** (ou **erreur de généralisation**) d'un prédicteur  $\hat{f}$  est défini par :

$$R(\hat{f}) = \mathbb{E} [\ell(\hat{f}(X), Y)].$$

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Oracle

Si on connaissait  $P_{X,Y}$ , on pourrait déterminer le prédicteur optimal, appelé **oracle** :

$$f^* = \arg \min_{f \in \mathcal{F}} R(f) .$$

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Exemples d'oracles

- ▶ Cas de la classification supervisée binaire :

Si  $\ell(y, y') = \mathbb{1}_{\{y \neq y'\}}$  alors :

$$g^*(x) = \begin{cases} 1 & \text{si } \mathbb{P}(Y = 1 / X = x) \geq \mathbb{P}(Y = -1 / X = x) \\ -1 & \text{sinon} \end{cases}$$

- ▶ Cas de la régression :

- ▶ Si  $\ell(y, y') = |y - y'|$  alors :

$$m^*(x) = \text{Med}(Y / X = x)$$

- ▶ Si  $\ell(y, y') = (y - y')^2$  alors :

$$m^*(x) = \mathbb{E}(Y / X = x)$$

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Enjeu

- ▶ L'objectif du data scientist est de déterminer une estimation  $\hat{f}$  de  $f$ , à partir de l'échantillon, telle que :

$$R(\hat{f}) \approx R(f^*) .$$

- ▶ En pratique, pour estimer  $f \in \mathcal{F}$  :
  1. On restreint  $\mathcal{F}$  à  $\mathcal{S}$ .
  2. On considère le risque empirique  $R_n$  (et non le risque).

D'où :

$$\hat{f} = \arg \min_{f \in \mathcal{S}} R_n(f) .$$

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Risque empirique

- ▶ Le risque empirique est défini par :

$$R_n(\hat{f}) = \frac{1}{n} \sum_{i=1}^n \ell(\hat{f}(X_i), Y_i).$$

- ▶ C'est un estimateur de  $R(\hat{f})$ .

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plan

Introduction

## Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

## Apprentissage supervisé

Formalisation

Pertes et risques

**Biais et variance**

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Biais et variance d'un estimateur I

- ▶ Dans le cas général ( $\mathcal{F}$ ), on cherche :

$$f^* = \arg \min_{f \in \mathcal{F}} R(f) .$$

- ▶ Dans le cas restreint ( $\mathcal{S} \subset \mathcal{F}$ ), on cherche :

$$f_S^* = \arg \min_{f \in \mathcal{S}} R(f) .$$

La décomposition biais (erreur d'approximation)-variance (erreur d'estimation) s'écrit :

$$R(\hat{f}_S) - R(f^*) = \underbrace{R(f_S^*) - R(f^*)}_{\text{erreur d'approximation}} + \underbrace{R(\hat{f}_S) - R(f_S^*)}_{\text{erreur d'estimation}} .$$

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

Biais et variance

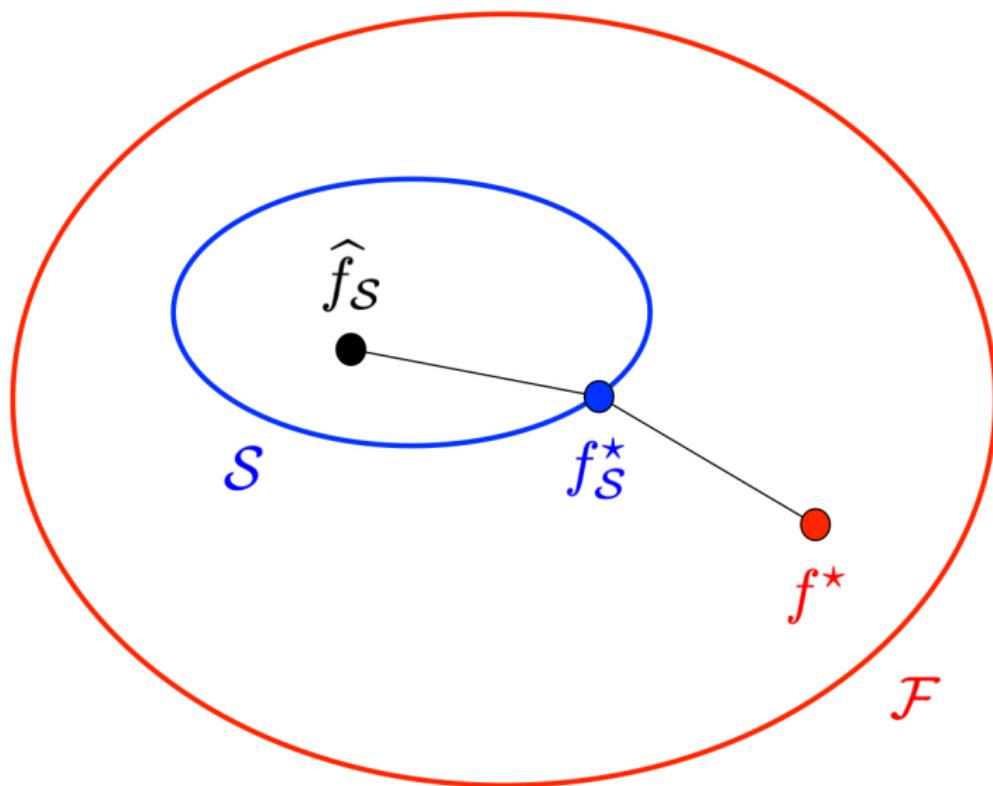
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Biais et variance d'un estimateur II



Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

**Biais et variance**

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Biais et variance d'un estimateur III

Introduction

Apprentissage  
supervisé

Formalisation  
Pertes et risques  
**Biais et variance**

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

Prévision de  
 $y \in \mathcal{Y}$

Echantillon 1

Prévision :  $\hat{y}^{(p)1} = \hat{f}_{d_n^{(1)}}(x)$

• • •

Echantillon  $K$

Prévision :  $\hat{y}^{(p)K} = \hat{f}_{d_n^{(K)}}(x)$

# Biais et variance d'un estimateur IV

Introduction

Apprentissage supervisé

Formalisation

Pertes et risques

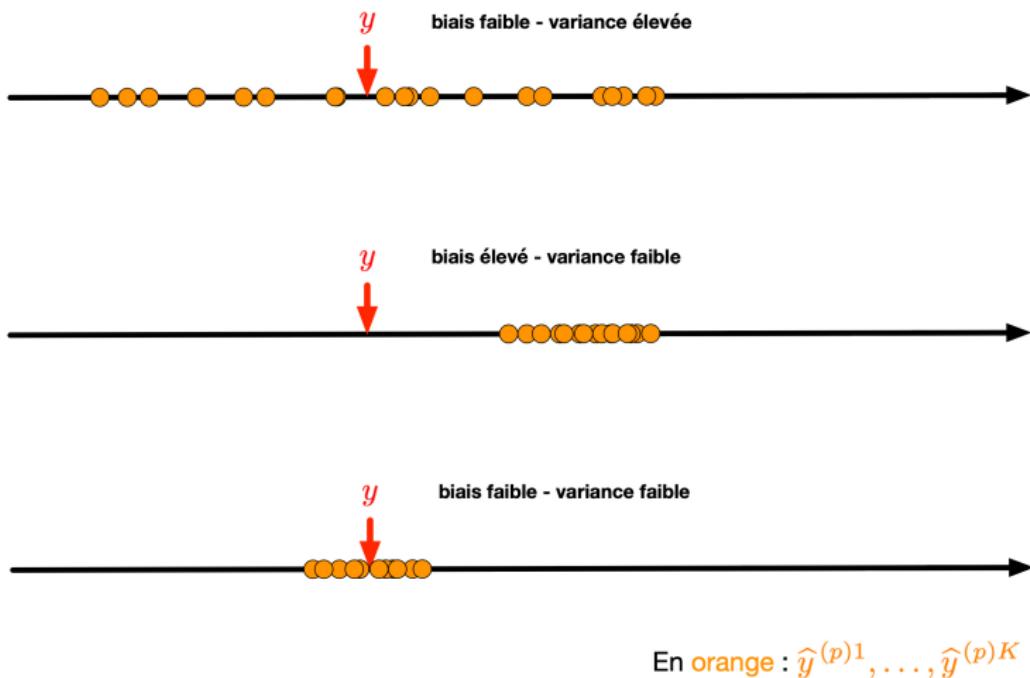
Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références



# Plan

Introduction

Apprentissage supervisé

**Sur-apprentissage**

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

**Sur-apprentissage**

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

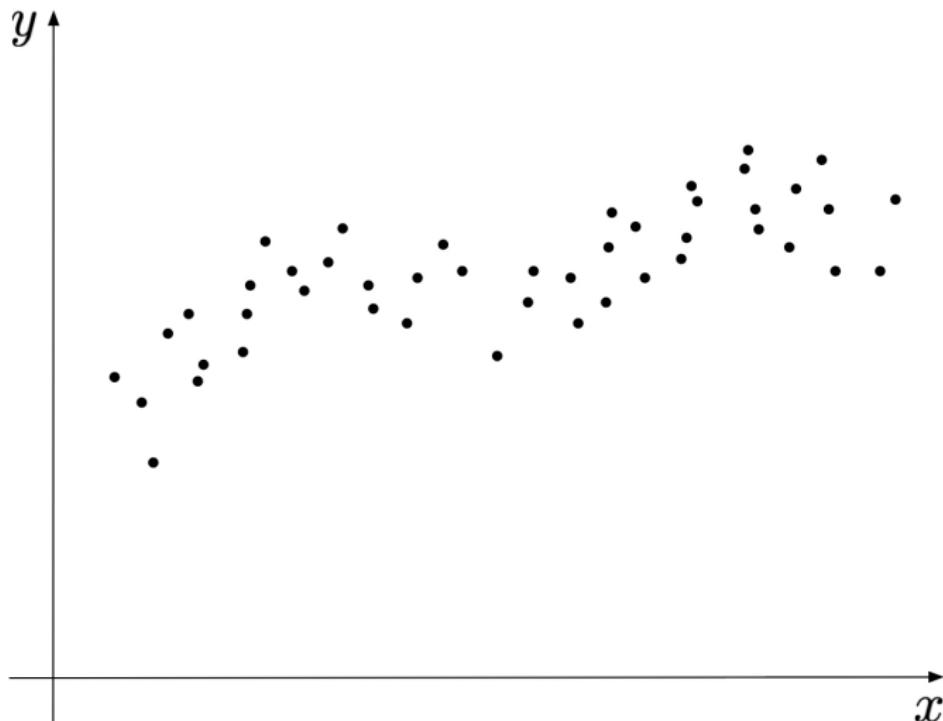
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Complexité I



Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

Problème

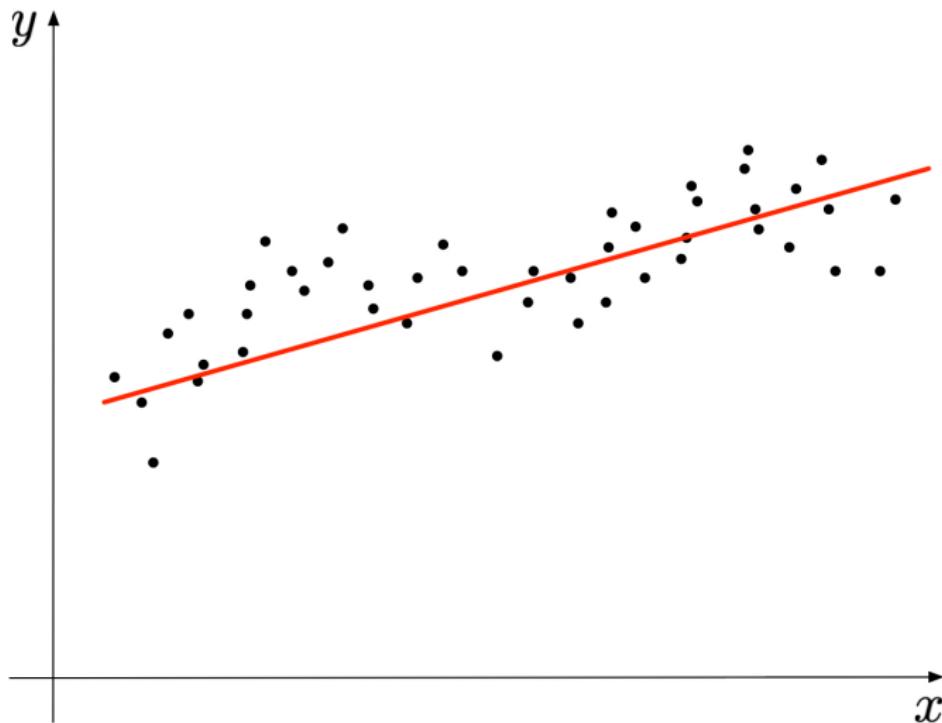
Palliatifs

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

# Complexité II



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

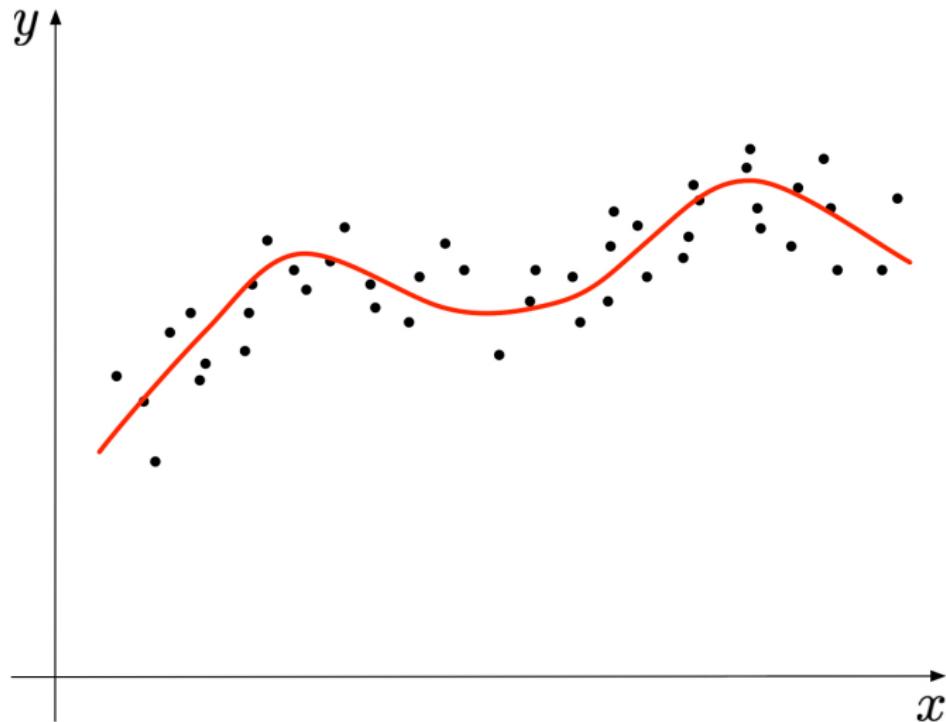
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Complexité III



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

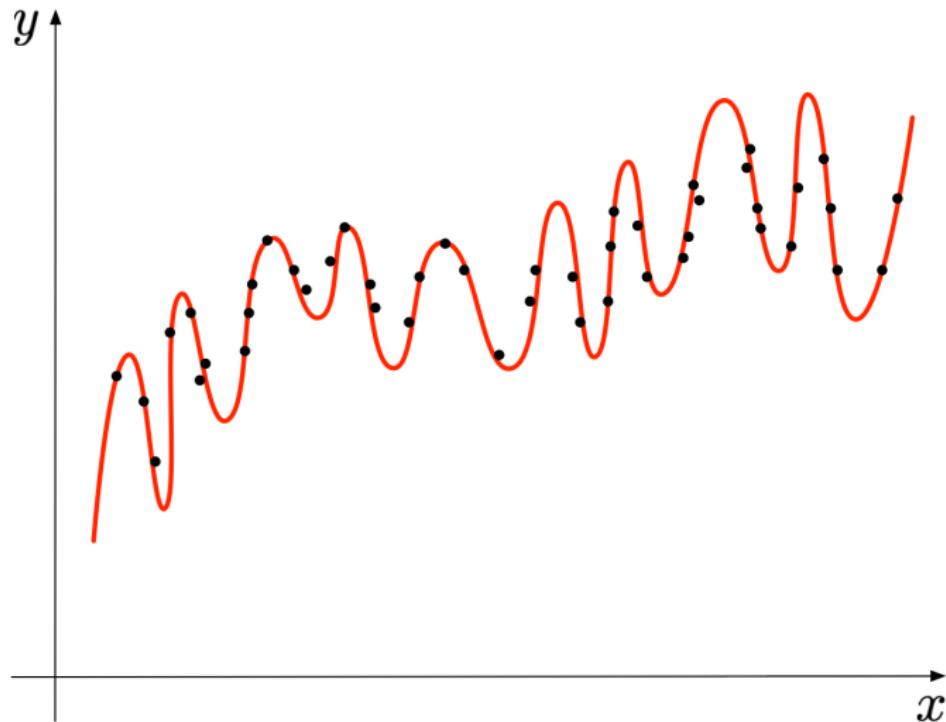
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Complexité IV



Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

**Problème**

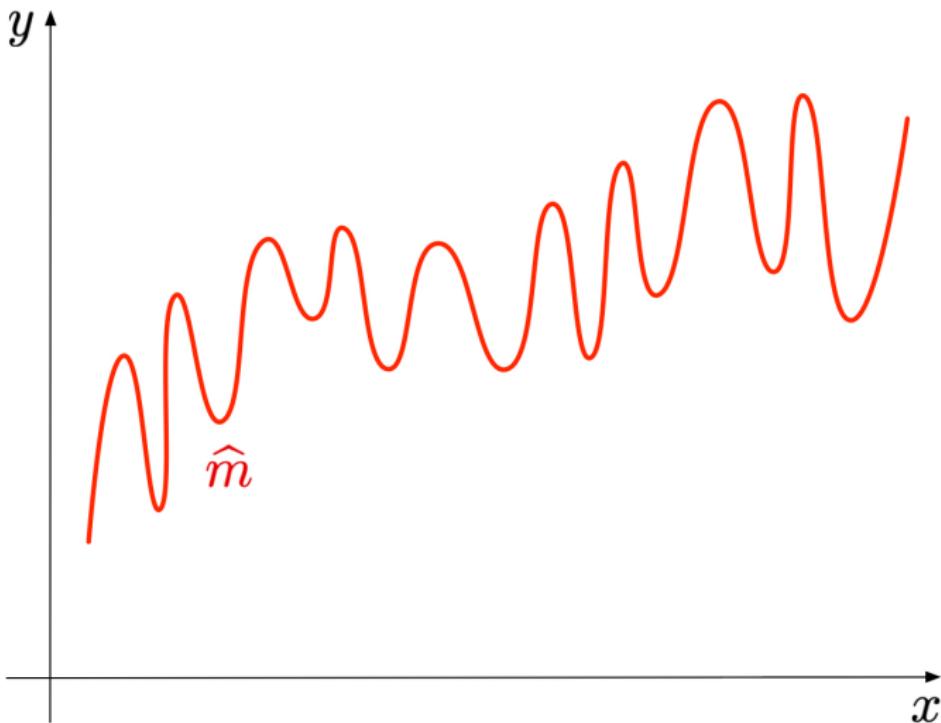
Palliatifs

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

# Complexité V



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

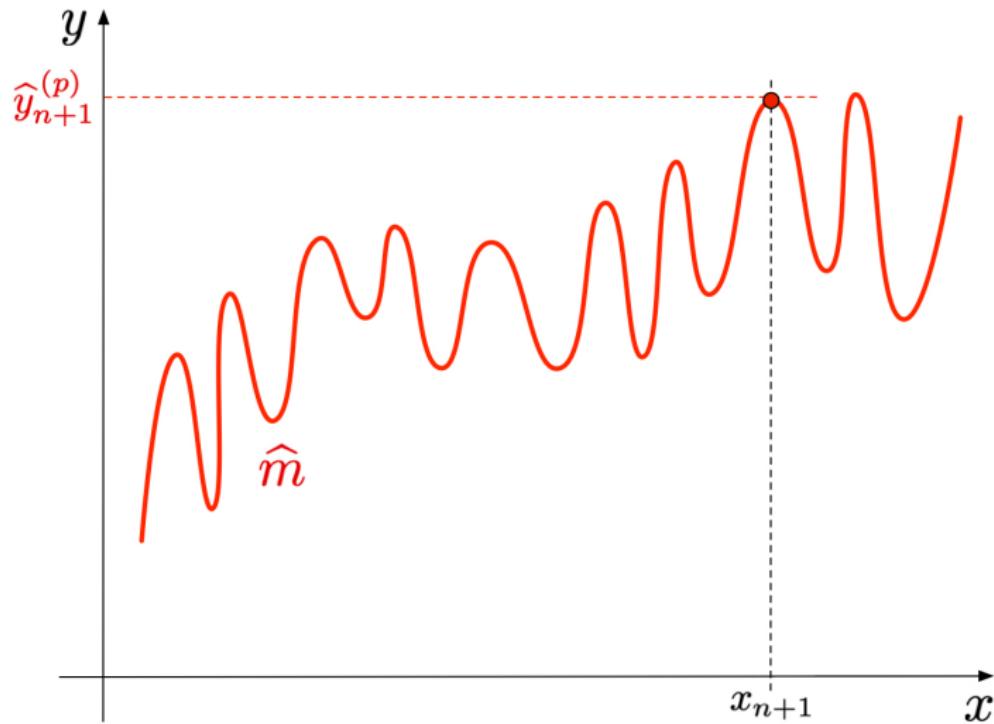
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Complexité VI



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

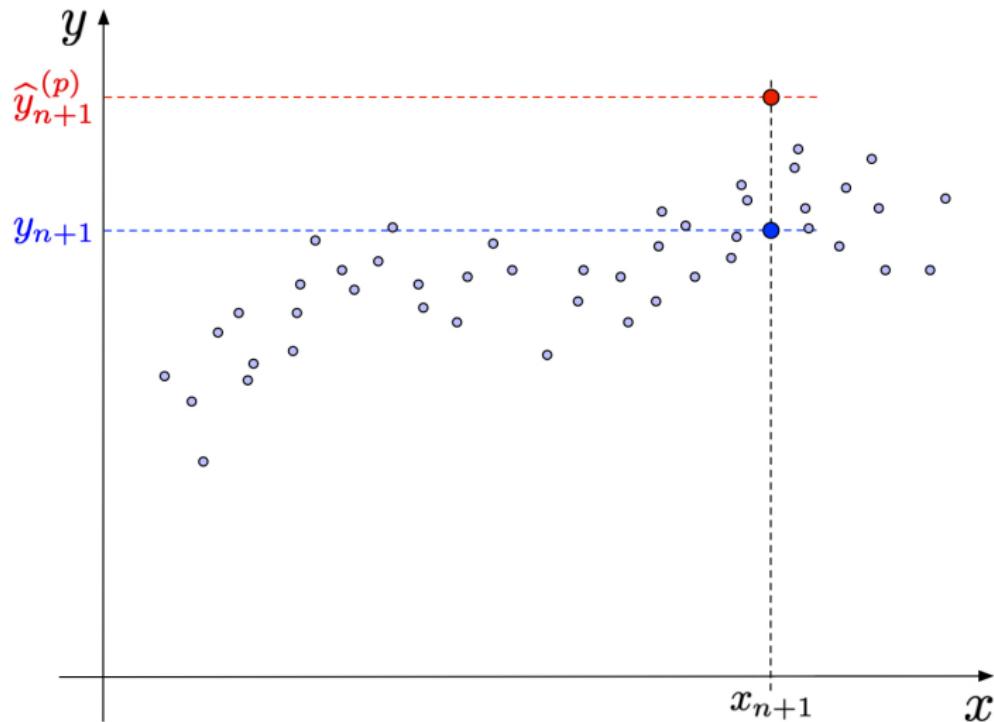
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Complexité VII



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

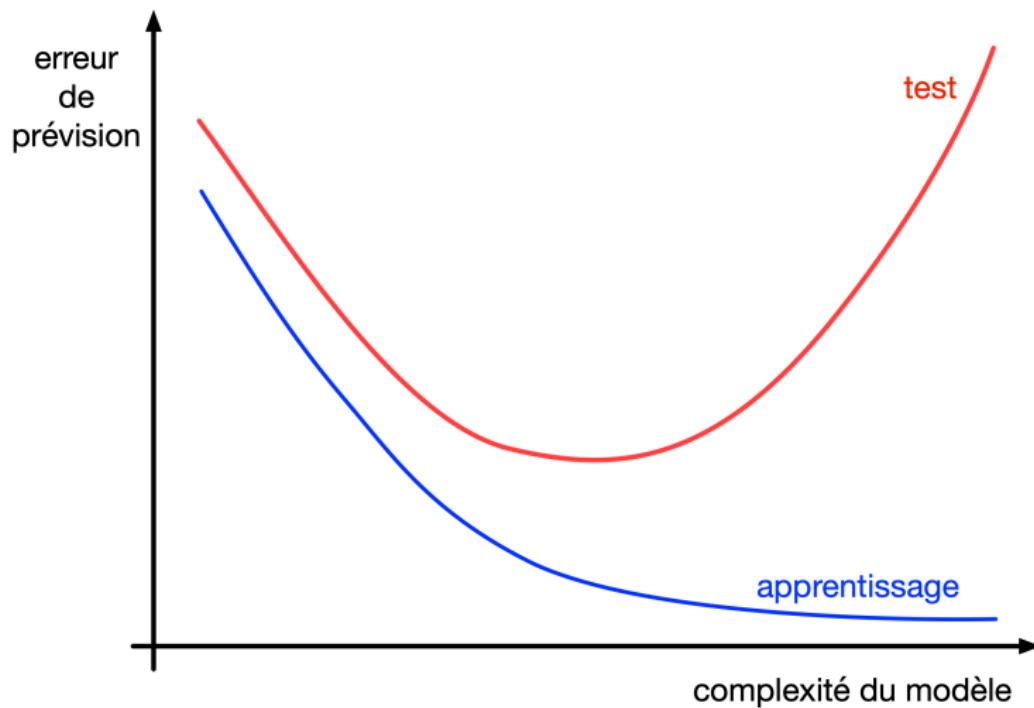
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

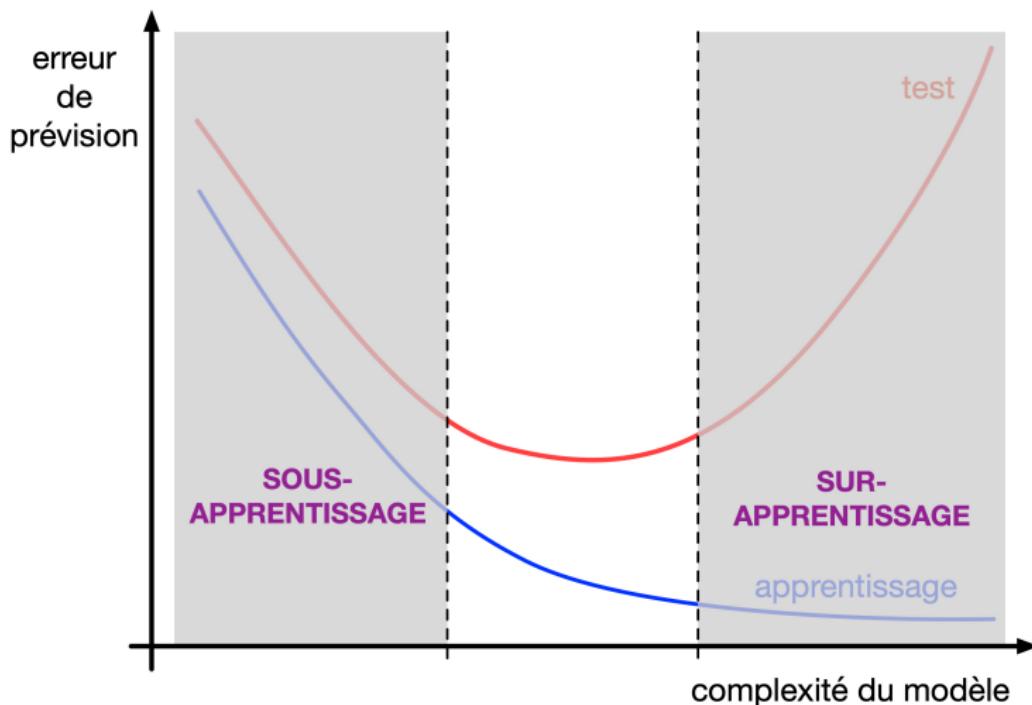
Références

# Erreur de prévision et complexité I



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Problème
- Palliatifs
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Références

# Erreur de prévision et complexité II



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

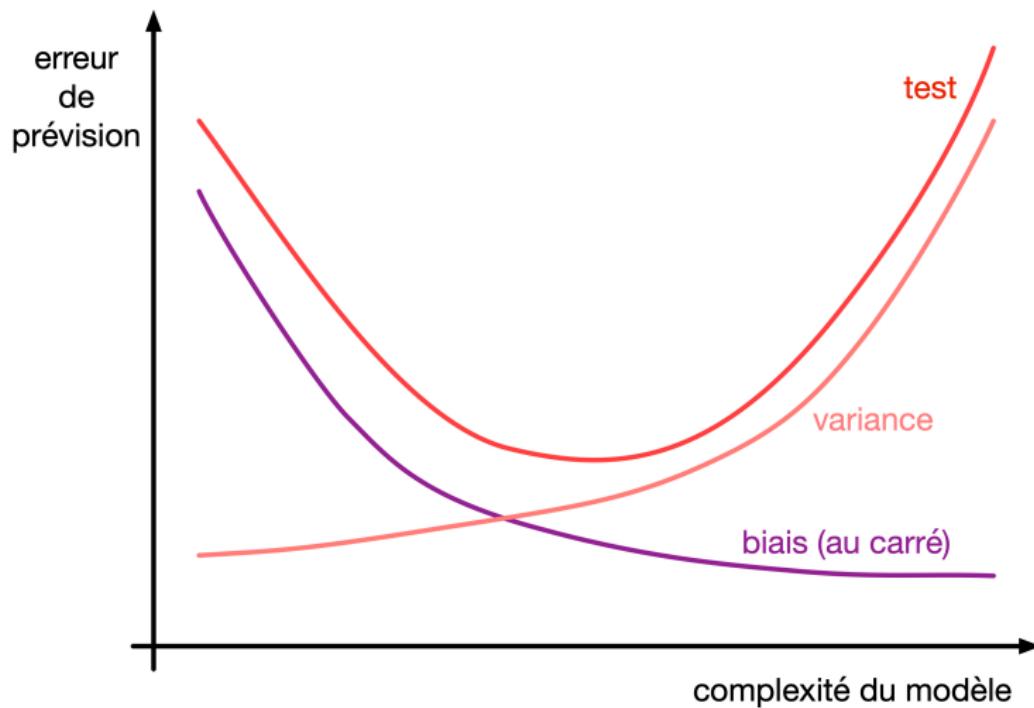
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Erreur de prévision et complexité III



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Plusieurs stratégies

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

- ▶ Utilisation de critères tels que l'AIC, le BIC, le  $C_p$  de Mallows... .
- ▶ Méthodes de rééchantillonage :
  - ▶ Apprentissage/test.
  - ▶ Validation croisée.

# Apprentissage/test : illustration I

Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

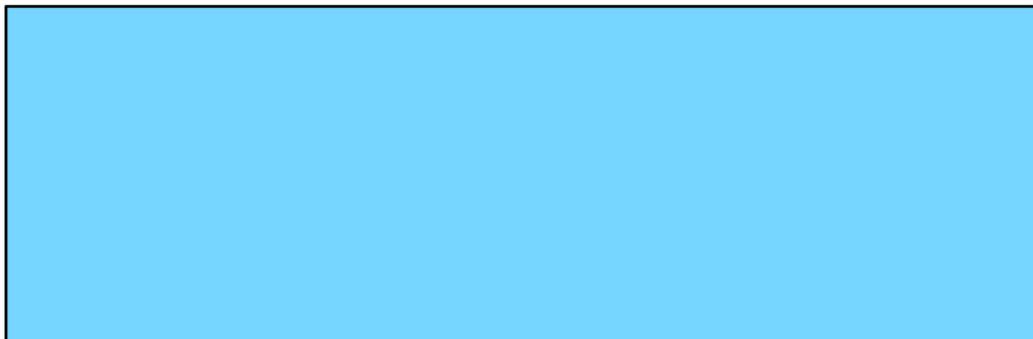
Problème

Palliatifs

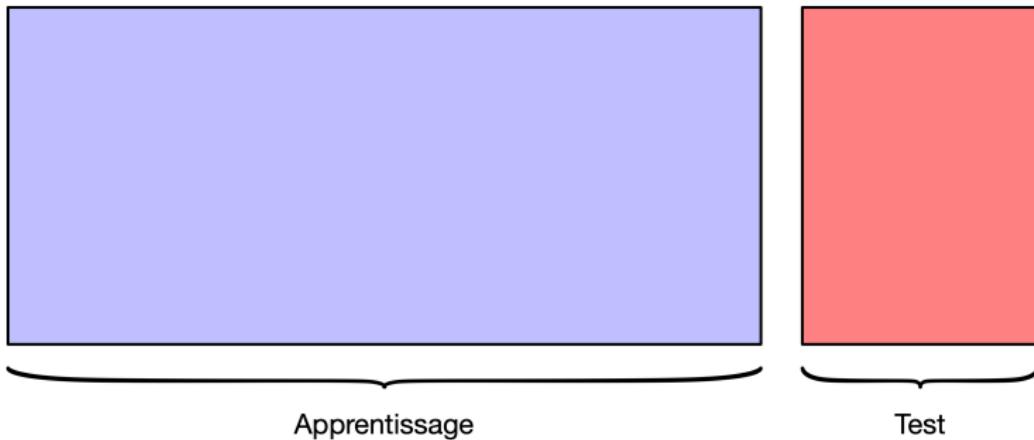
Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références



# Apprentissage/test : illustration II



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

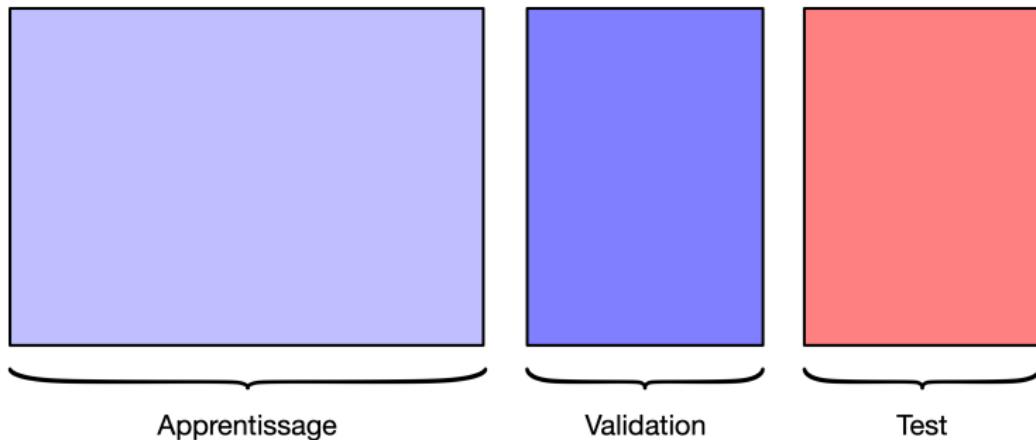
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Apprentissage/test : illustration III



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : principe

1. Diviser aléatoirement les données en  **$K$  blocs** (égaux ou équivalents).

Le bloc  $k$  contient  $n_k$  observations :  $n_k = \frac{n}{K}$  si  $n$  est un multiple de  $K$ .

2. Pour  $k \in \{1, \dots, K\}$  :

- 2.1 Retirer le bloc  $k$  de la base d'apprentissage.
- 2.2 Estimer la fonction de prévision sur la base d'apprentissage.
- 2.3 Calculer un critère d'erreur de prévision sur le bloc  $k$  :  **$CV_k$**  (ex : MSE pour la régression).

3. Calculer le critère de validation croisée :

$$CV = \sum_{k=1}^K \frac{n_k}{n} CV_k .$$

Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

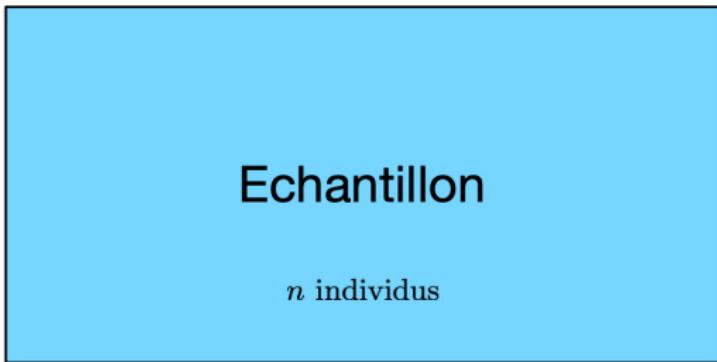
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration I



Introduction

Apprentissage  
supervisé

Sur-  
apprentissage  
Problème  
Palliatifs

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

# Validation croisée : illustration II

Bloc 1  
 $n_1$

Bloc 2  
 $n_2$

Bloc 3  
 $n_3$

Bloc 4  
 $n_4$

Bloc 5  
 $n_5$

Introduction

Apprentissage  
supervisé

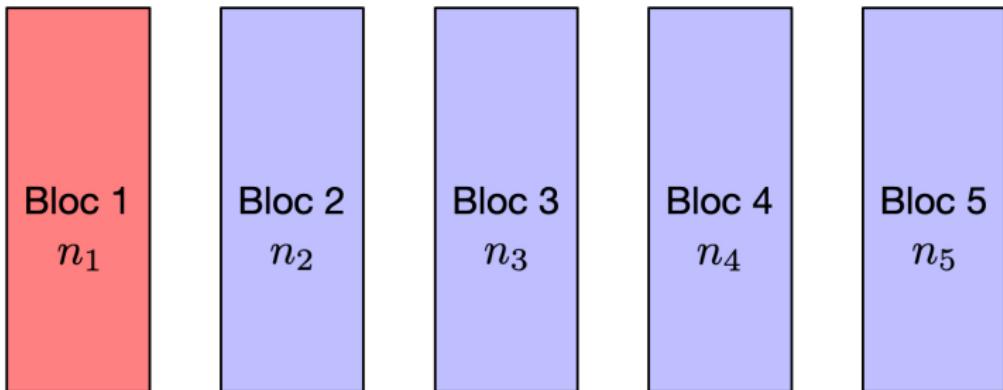
Sur-  
apprentissage  
Problème  
Palliatifs

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

# Validation croisée : illustration III



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

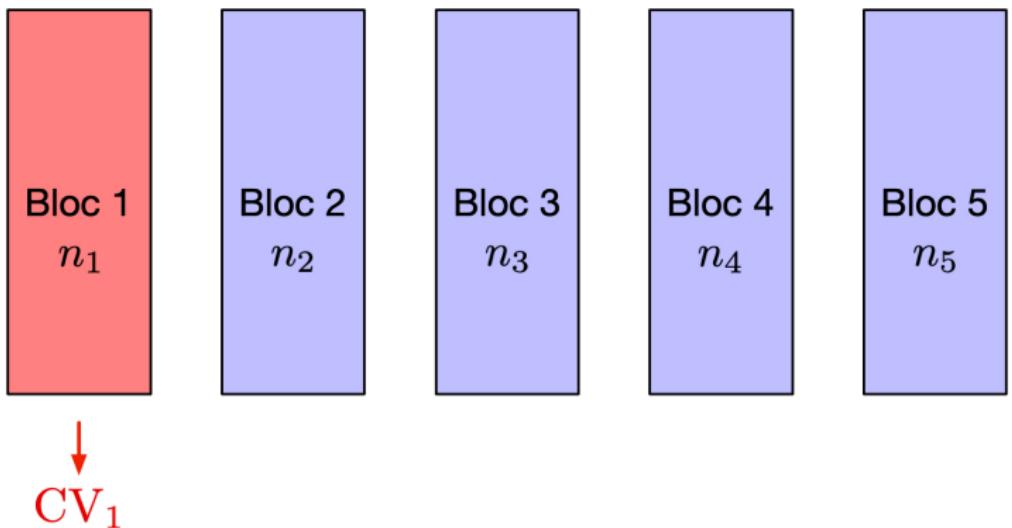
Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

- █ Apprentissage
- █ Test

# Validation croisée : illustration IV



■ Apprentissage  
■ Test

Introduction

Apprentissage supervisé

Sur-apprentissage

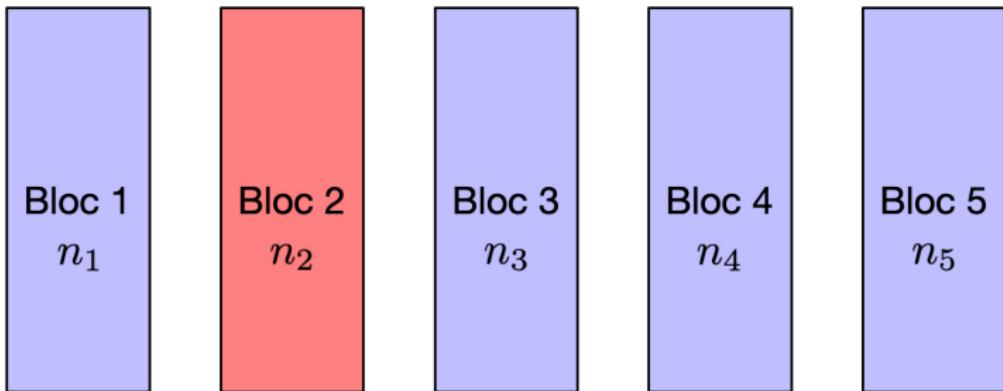
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration V



- █ Apprentissage
- █ Test

Introduction

Apprentissage supervisé

Sur-apprentissage

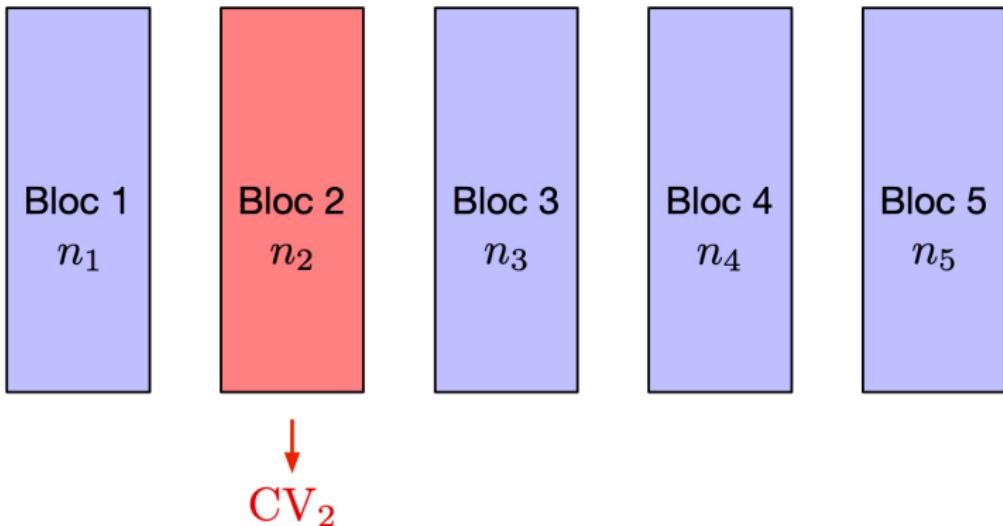
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration VI



- Apprentissage
- Test

Introduction

Apprentissage supervisé

Sur-apprentissage

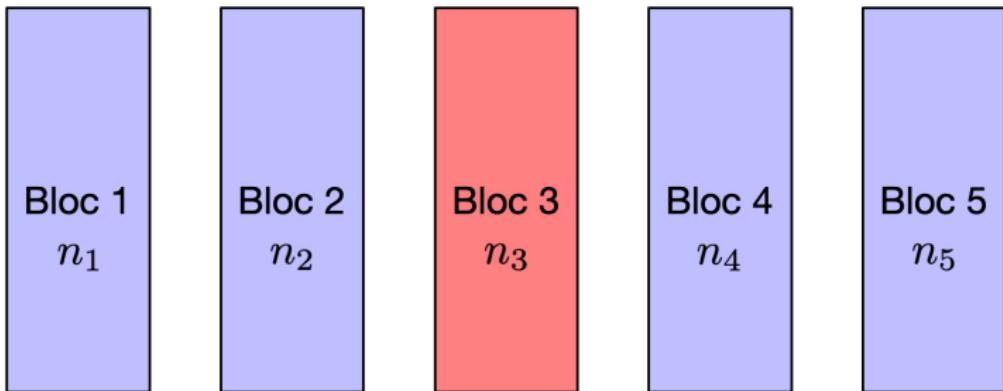
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration VII



- █ Apprentissage
- █ Test

Introduction

Apprentissage supervisé

Sur-apprentissage

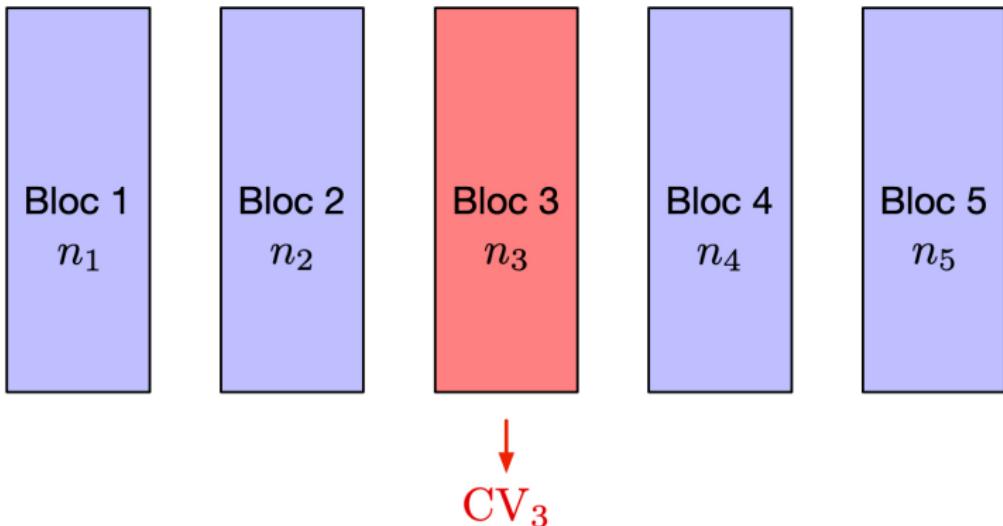
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration VIII



■ Apprentissage  
■ Test

Introduction

Apprentissage supervisé

Sur-apprentissage

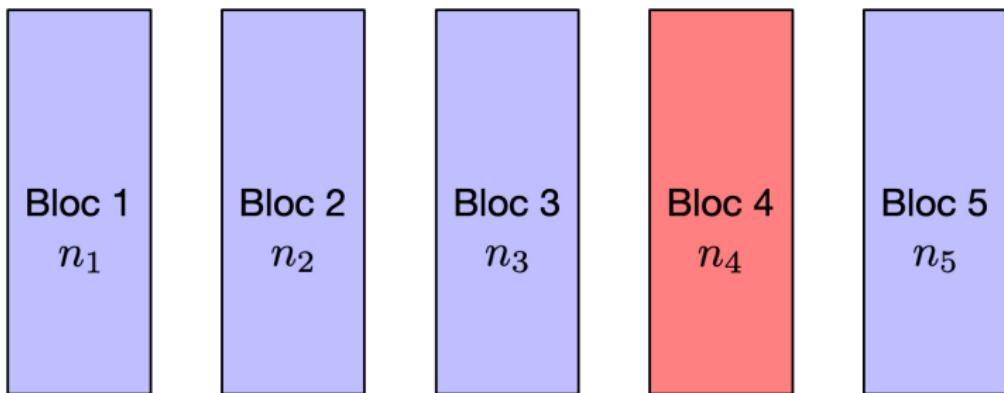
Problème Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration IX



■ Apprentissage  
■ Test

Introduction

Apprentissage supervisé

Sur-apprentissage

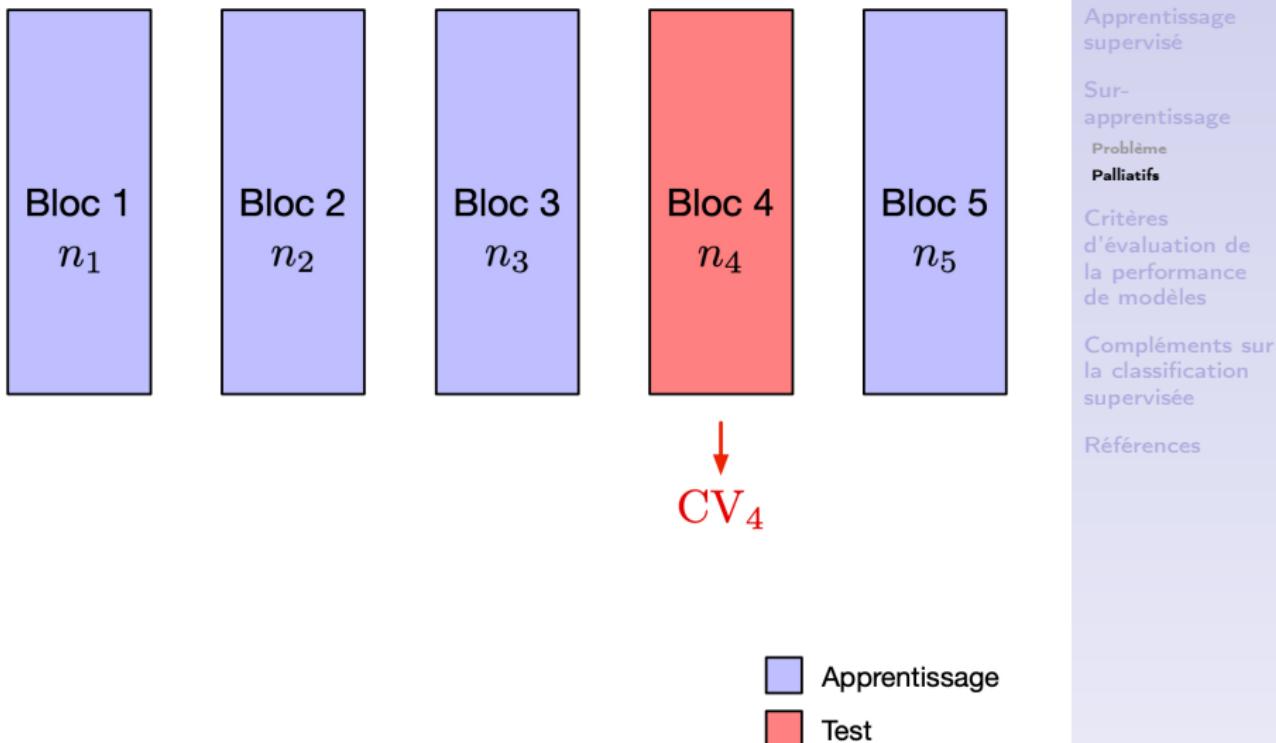
Problème Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration X



Introduction

Apprentissage supervisé

Sur-apprentissage

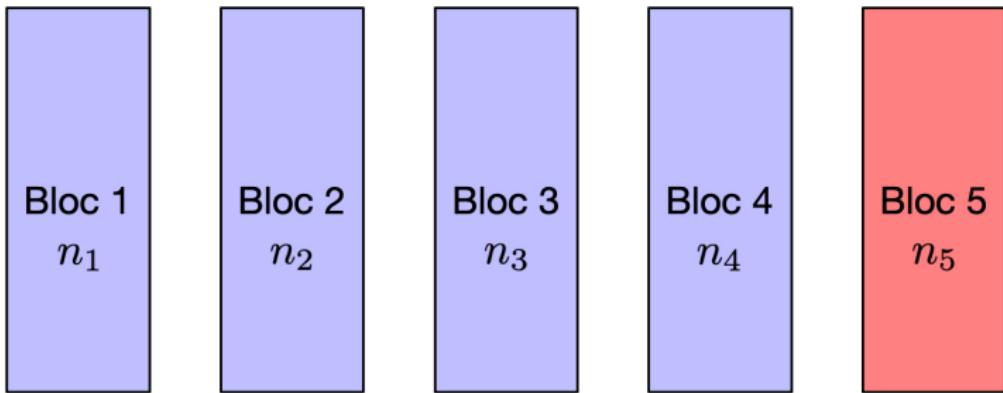
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée : illustration XI



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème

Palliatifs

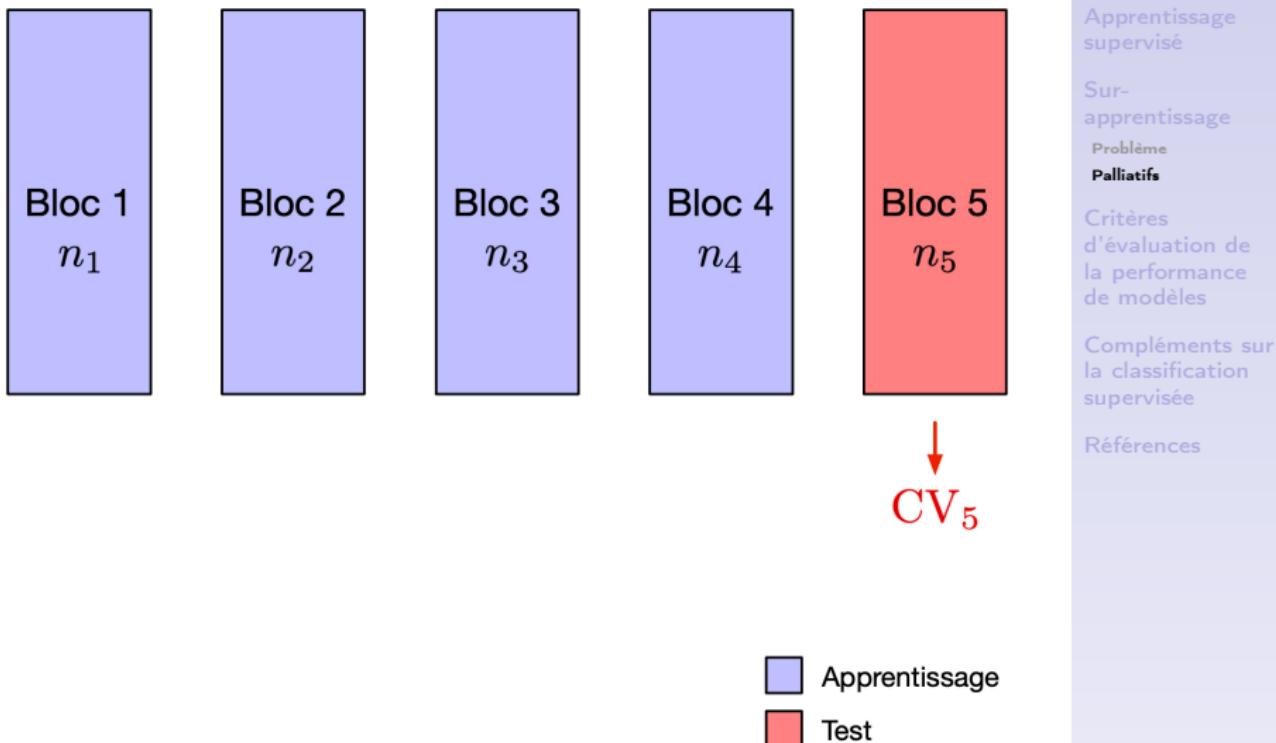
Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

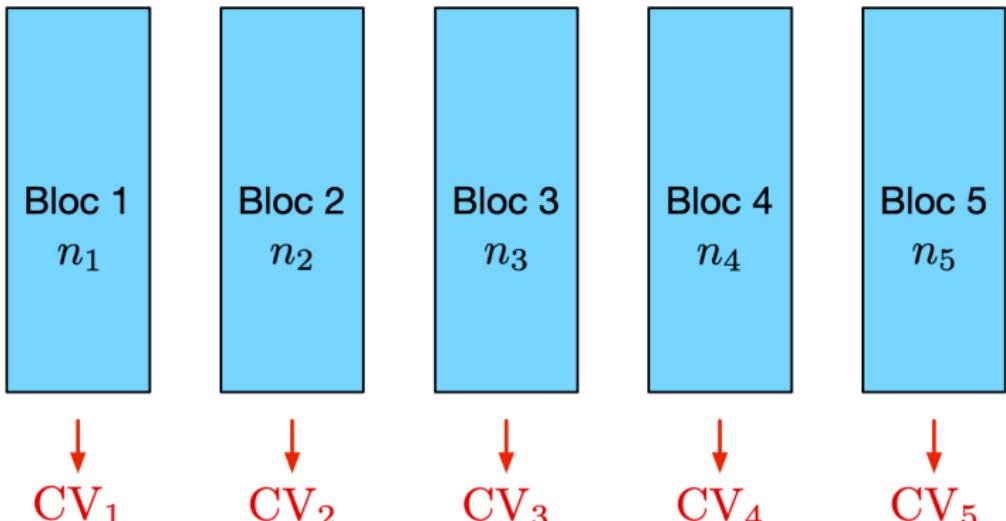
Références

- █ Apprentissage
- █ Test

# Validation croisée : illustration XII



## Validation croisée : illustration XIII



$$CV = \sum_{k=1}^5 \frac{n_k}{n} CV_k$$

Introduction

Apprentissage supervisé

Sur-apprentissage

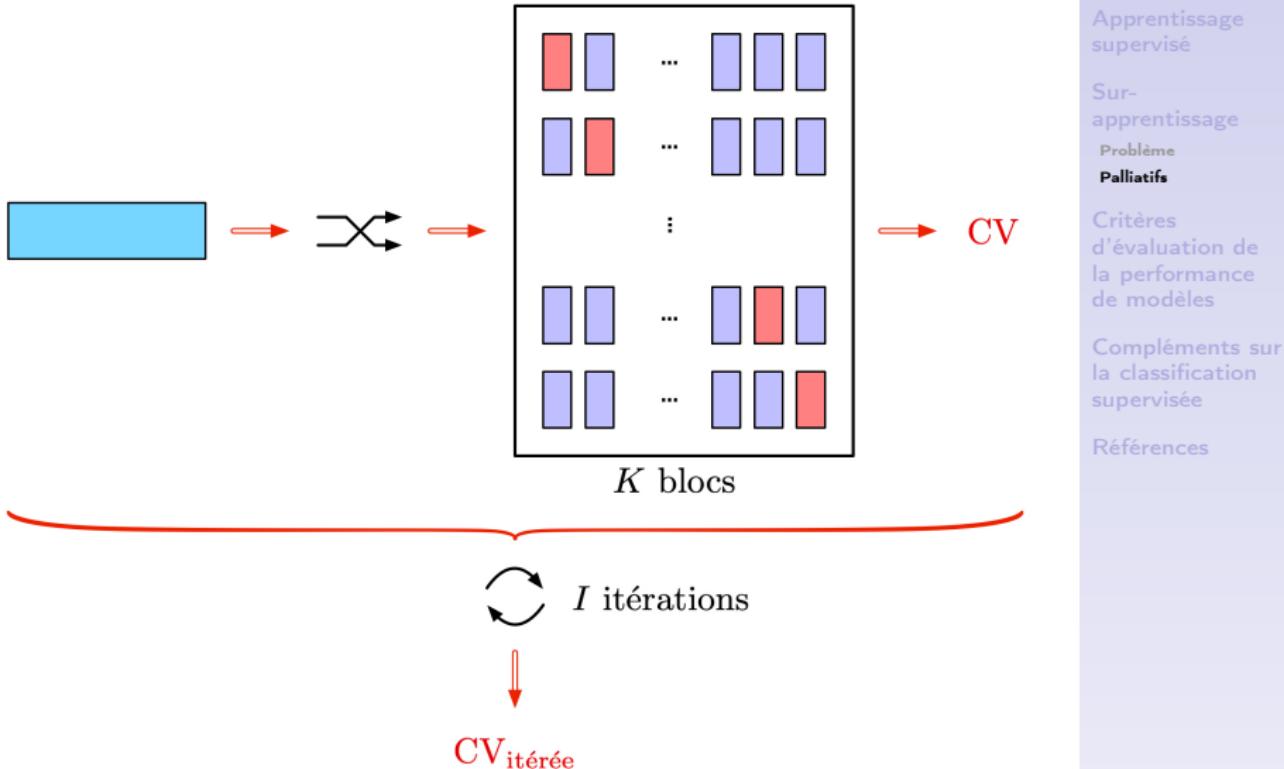
Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Validation croisée itérée



Introduction

Apprentissage supervisé

Sur-apprentissage

Problème  
Palliatifs

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

# Remarques

Introduction

Apprentissage  
supervisé

Sur-  
apprentissage  
Problème  
Palliatifs

Critères  
d'évaluation de  
la performance  
de modèles

Compléments sur  
la classification  
supervisée

Références

- ▶ Usuellement :  $K = 5$  ou  $K = 10$ .
- ▶ Lorsque  $K = n$  : on parle d'estimateur « **leave one out** » (LOO)

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

**Critères d'évaluation de la performance de modèles**

Régression

Classification supervisée

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Sur-apprentissage

**Critères d'évaluation de la performance de modèles**

Régression

Classification supervisée

Compléments sur la classification supervisée

Références

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

**Régression**

Classification supervisée

Compléments sur la classification supervisée

Références

# RMSE et nRMSE

Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Régression

Classification  
supervisée

Compléments sur  
la classification  
supervisée

Références

- Le **RMSE** (Root Mean Square Error) vaut :

$$\text{RMSE} = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_i - y_i)^2}.$$

- Le **nRMSE** (normalized RMSE) vaut :

$$\text{nRMSE} = \frac{\text{RMSE}}{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \hat{y}_i}.$$

# MAE et MAPE

Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Régression

Classification  
supervisée

Compléments sur  
la classification  
supervisée

Références

- Le **MAE** (Mean Absolute Error) vaut :

$$\text{MAE} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} |\hat{y}_i - y_i| .$$

- Le **MAPE** (Mean Absolute Percent Error) vaut :

$$\text{MAPE} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100 .$$

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

**Classification supervisée**

Compléments sur la classification supervisée

Références

# Matrice de confusion

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Références

- ▶ Dans le cas de la classification supervisée binaire, la matrice de confusion vaut :

		Prévision	
		1 (Positif)	0 (Négatif)
Vérité	1 (Positif)	Vrai positif (VP)	Faux négatif (FN)
	0 (Négatif)	Faux positif (FP)	Vrai négatif (VN)

- ▶ Dans le cas de la classification supervisée multi-classes, on peut établir la matrice de confusion, avec autant de lignes et de colonnes que de classes, et en déduire les nombres VP, FP, VN et FN.

# Critères de qualité

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

**Classification supervisée**

Compléments sur la classification supervisée

Références

- ▶ On considère usuellement :

- ▶ L'exactitude.
- ▶ La spécificité.
- ▶ La précision.
- ▶ La sensibilité
- ▶ Le  $F_1$ .
- ▶ L'AUC.

- ▶ Ces indicateurs prennent leurs valeurs sur  $[0, 1]$  : plus ils sont proches de 1, meilleur est le modèle.

# Exactitude (et erreur de classification) et spécificité

- ▶ L'**exactitude** (*accuracy*) vaut :

$$\text{exactitude} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{VN} + \text{FN}}.$$

Notons que l'erreur de classification (*classification error*) vaut :

$$\text{erreur} = \frac{\text{FP} + \text{FN}}{\text{VP} + \text{FP} + \text{VN} + \text{FN}} = 1 - \text{exactitude}.$$

- ▶ La **spécificité** (*specificity*), le taux de négatifs classés négatifs (« vrais négatifs »), vaut :

$$\text{spécificité} = \frac{\text{VN}}{\text{FP} + \text{VN}}.$$

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Références

# Précision, sensibilité et $F_1$

- ▶ La **précision** (*precision*), ou valeur prédictive positive, vaut :

$$\text{précision} = \frac{\text{VP}}{\text{VP} + \text{FP}}.$$

- ▶ La **sensibilité** (*sensitivity*), ou rappel (*recall*), est le taux de positifs classés positifs (« vrais positifs ») :

$$\text{sensibilité} = \frac{\text{VP}}{\text{VP} + \text{FN}}.$$

- ▶ Le score  $F_1$  est la moyenne harmonique de la précision et de la sensibilité :

$$\begin{aligned} F_1 &= \frac{2}{\frac{1}{\text{précision}} + \frac{1}{\text{sensibilité}}} \\ &= 2 \frac{\text{précision} \cdot \text{sensibilité}}{\text{précision} + \text{sensibilité}}. \end{aligned}$$

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Références

# Courbe ROC I

- ▶ La courbe **ROC** (Receiver Operating Characteristic) représente la sensibilité (taux de vrais positifs) en fonction de l'anti-spécificité (taux de faux positifs) pour différents seuils de décision  $s$  :

$$\hat{y}_i = 1 \quad \text{si} \quad \mathbb{P}(Y = 1 / X_1 = x_{i1}, \dots, X_p = x_{ip}) > s .$$

- ▶ Plus le seuil  $s$  est important :
  - ▶ plus le taux de vrais positifs est important,
  - ▶ moins le taux de faux positifs est important.
- ▶ La courbe ROC est croissante et au-dessus de la première bissectrice (correspondant à une prédiction de type « tirage au sort »).
- ▶ La prédiction « optimale » fournirait une courbe ROC égale à 0 pour  $s = 0$  et égale à 1 pour  $s \in ]0, 1]$ .

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

Classification supervisée

Compléments sur la classification supervisée

Références

# Courbe ROC II

Introduction

Apprentissage supervisé

Sur-apprentissage

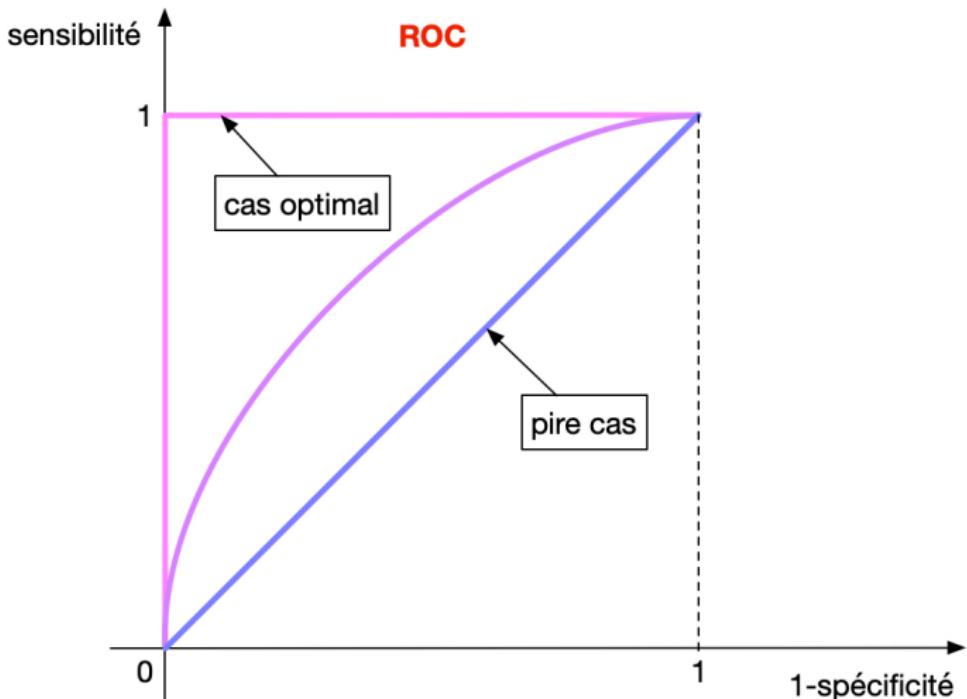
Critères d'évaluation de la performance de modèles

Régression

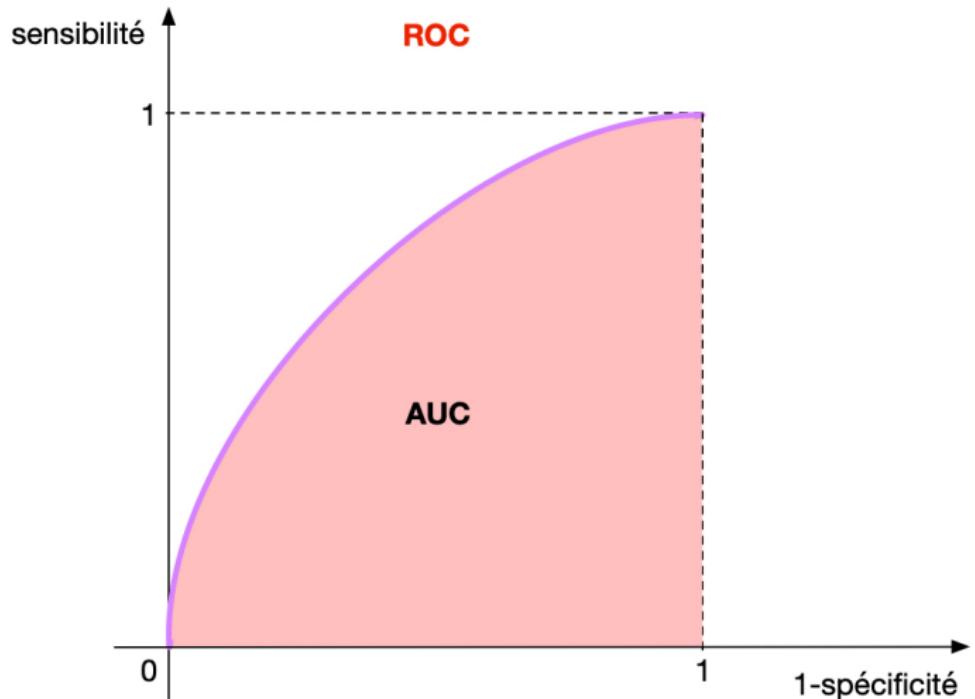
**Classification supervisée**

Compléments sur la classification supervisée

Références



# AUC I



Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Régression

**Classification supervisée**

Compléments sur la classification supervisée

Références

## AUC II

Introduction

Apprentissage  
supervisé

Sur-  
apprentissage

Critères  
d'évaluation de  
la performance  
de modèles

Régression

Classification  
supervisée

Compléments sur  
la classification  
supervisée

Références

L'aire sous la courbe ROC, l'**AUC** (Area Under the ROC), est une mesure de la qualité de la classification et varie entre :

- ▶  $\text{AUC} = \frac{1}{2}$  : le pire des cas (prédiction de type « tirage au sort »),
- ▶  $\text{AUC} = 1$  : le meilleur des cas (prédiction « optimale »).

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

**Compléments sur la classification supervisée**

Classification supervisée multi-classes

Données déséquilibrées

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

**Compléments sur la classification supervisée**

Classification supervisée multi-classes

Données déséquilibrées

Références

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

**Compléments sur la classification supervisée**

Classification supervisée multi-classes

Données déséquilibrées

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# Pour des méthodes de classification supervisée binaire

Si les méthodes de classification supervisée n'intègrent que le cas binaire (ex : SVM), il existe alors différentes stratégies dans le cas où on a  $K > 2$  classes, parmi lesquelles :

- ▶ Stratégie « une contre toutes » **OvR** (One vs Rest)
  1. On effectue les  $K$  classifications supervisées binaires  $Y = k$  contre  $Y \neq k$  pour  $k \in \{1, \dots, K\}$ .
  2. On affecte à une observation la classe qui a la probabilité la plus élevée parmi ces  $K$  classifications supervisées binaires.
- ▶ Stratégie « une contre une » **OvO** (One vs One)
  1. On effectue les  $\binom{K}{2}$  classifications supervisées binaires  $Y = k$  contre  $Y = k'$  pour  $(k, k') \in \{1, \dots, K\}^2$  et  $k \neq k'$ .
  2. On affecte à une observation la classe majoritaire parmi ces  $\binom{K}{2}$  classifications supervisées binaires.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# Plan

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

**Compléments sur la classification supervisée**

Classification supervisée multi-classes

Données déséquilibrées

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# Problématique

- ▶ Des données déséquilibrées peuvent conduire à une évaluation fallacieuse des modèles.
- ▶ La notion de déséquilibre n'est pas définie formellement mais un ratio de 1 à 10 entre 2 classes est un ordre de grandeur communément considéré.
- ▶ Le déséquilibre sera plus fortement ressenti lorsque la taille de l'échantillon est petite.
- ▶ En tout premier lieu, il faut évaluer l'opportunité de disposer de nouvelles données et toujours adopter les critères de mesure les plus adéquats pour la qualité des modèles.
- ▶ On utilise classiquement des méthodes de rééchantillonnage.
- ▶ On utilise ces algorithmes uniquement sur l'échantillon d'apprentissage, pas sur ceux de validation ou de test.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# Sur-échantillonnage

On peut procéder à du **sur-échantillonnage** (*over-sampling*) :

- ▶ Méthode naïve : dupliquer aléatoirement des observations de la classe minoritaire jusqu'à l'obtention du ratio classe minoritaire-classe majoritaire souhaité.
- ▶ Un des algorithmes les plus utilisés est **SMOTE** (*Synthetic Minority Over-sampling TErchnique*) pour des covariables quantitatives. Pour des covariables qualitatives, il existe une alternative : SMOTENC (SMOTE Nominal Continous).  
Référence : [\(Chawla et collab., 2002\)](#)
- ▶ Il existe des variantes à SMOTE comme **BorderSMOTE** et **ADASYN** (ADAptive SYNthetic) qui permettent de pallier la génération d'observations « dangereuses », près des frontières entre les classes. ADASYN prend en considération la densité des observations de la classe minoritaire.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# SMOTE I

Pour chaque **observation  $i$  de la classe minoritaire** :

1. Déterminer ses  **$k$  plus proches voisins** ( $k$  fixé, inférieur au nombre d'observations de la classe minoritaire).
2. **Tirer au sort** un de ses  $k$  plus proches voisins  $i'$ .
3. Calculer les **écart entre toutes les covariables** des individus  $i$  et  $i'$  (dans le cas où les covariables sont quantitatives).
4. Multiplier ces écarts par un **nombre aléatoire sur  $[0, 1]$** .
5. Ajouter ces valeurs aux covariables de l'individu  $i$  afin de créer **un nouvel individu de la classe minoritaire**.

**Répéter** cette opération autant de fois que nécessaire pour obtenir le ratio classe minoritaire-classe majoritaire souhaité (l'algorithme originel inclut en sus un sous-échantillonnage aléatoire parmi les observations de la classe majoritaire).

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

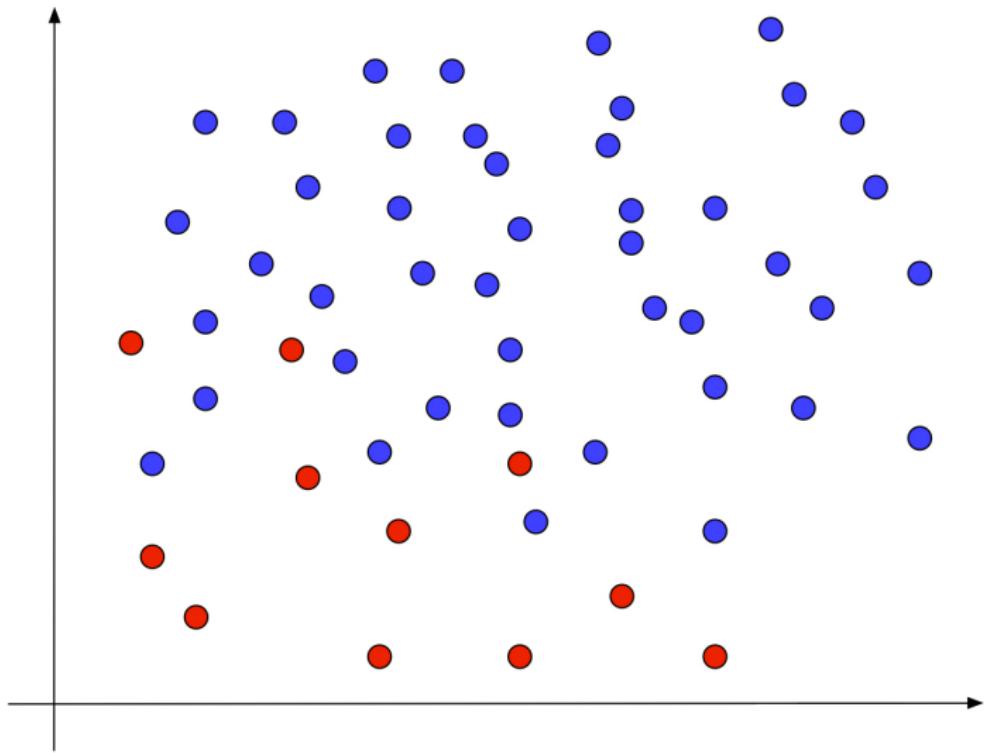
Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

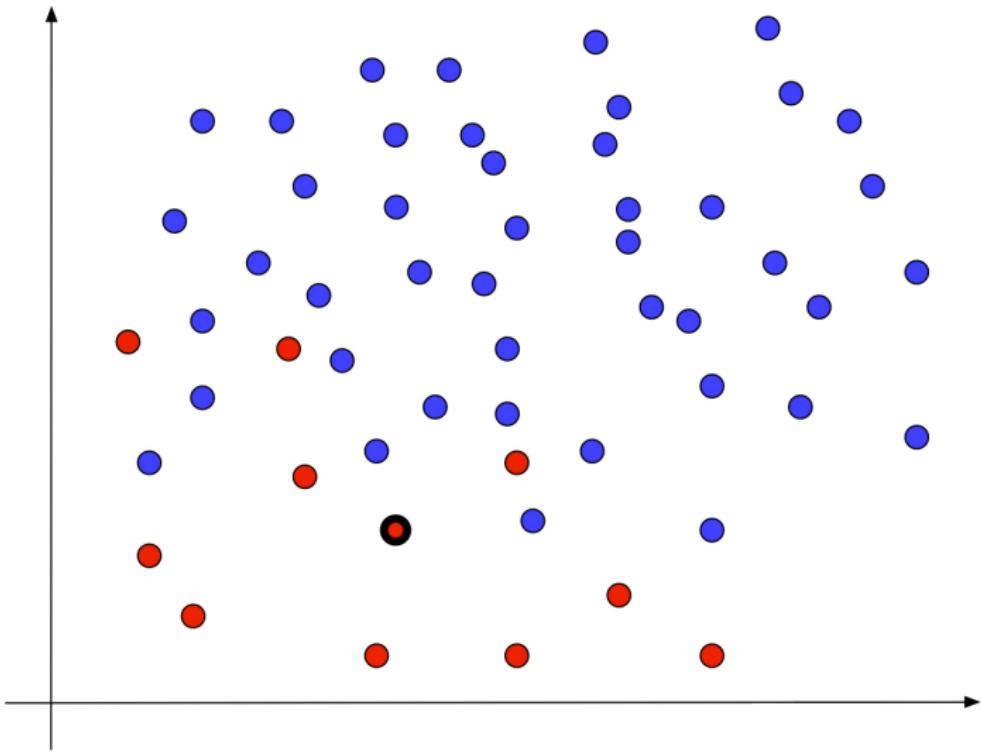
# SMOTE II



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

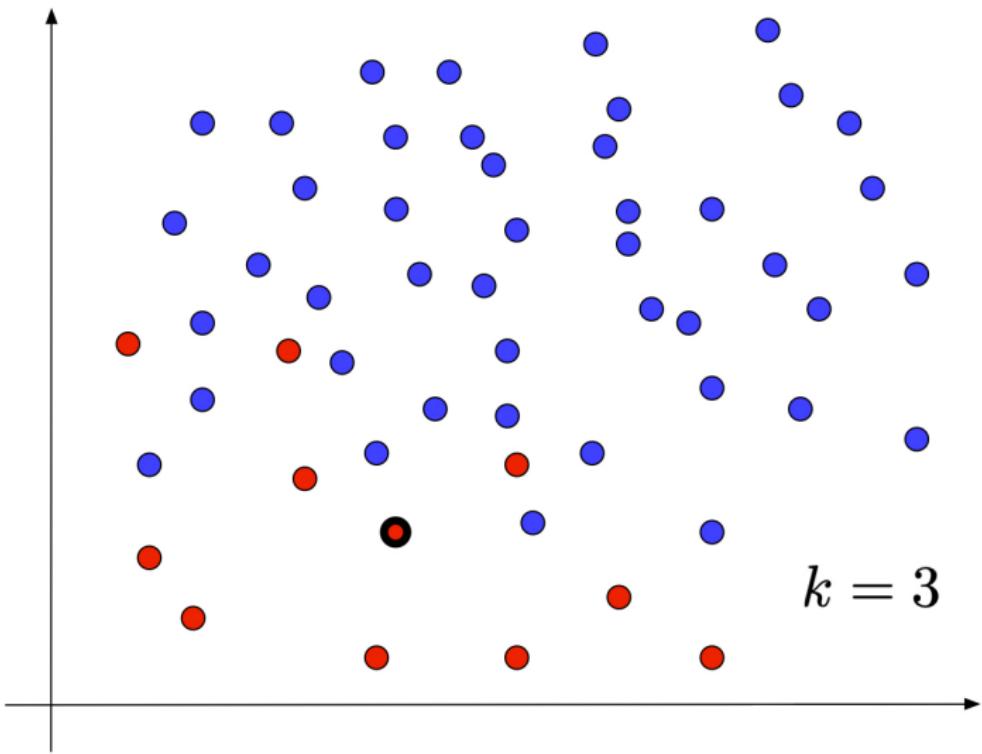
# SMOTE III



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

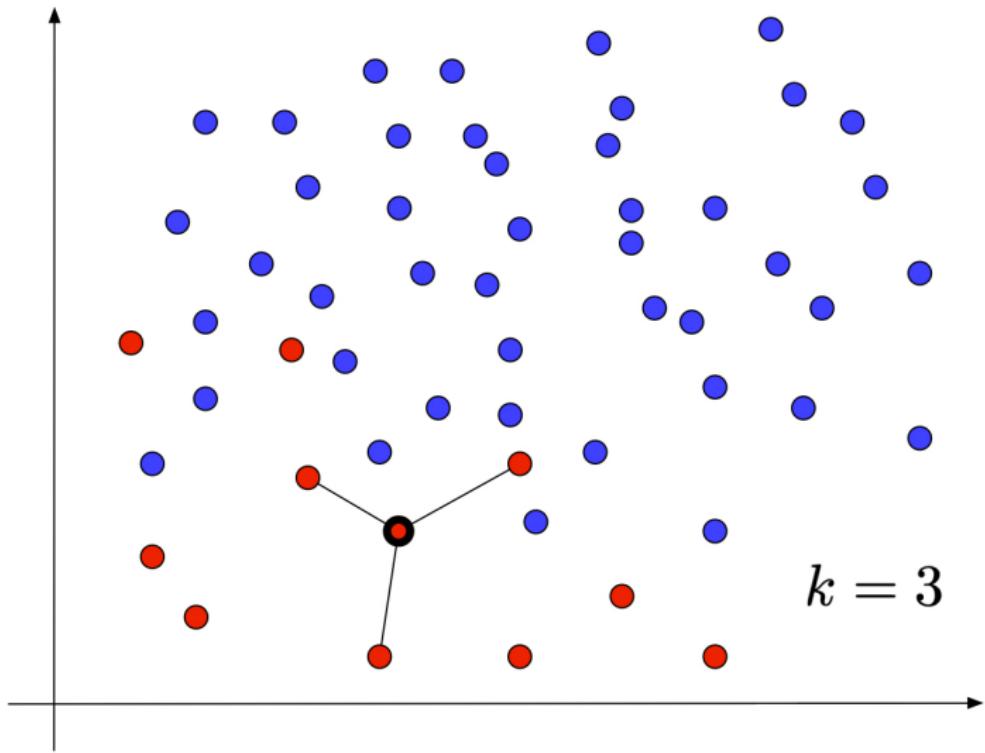
# SMOTE IV



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

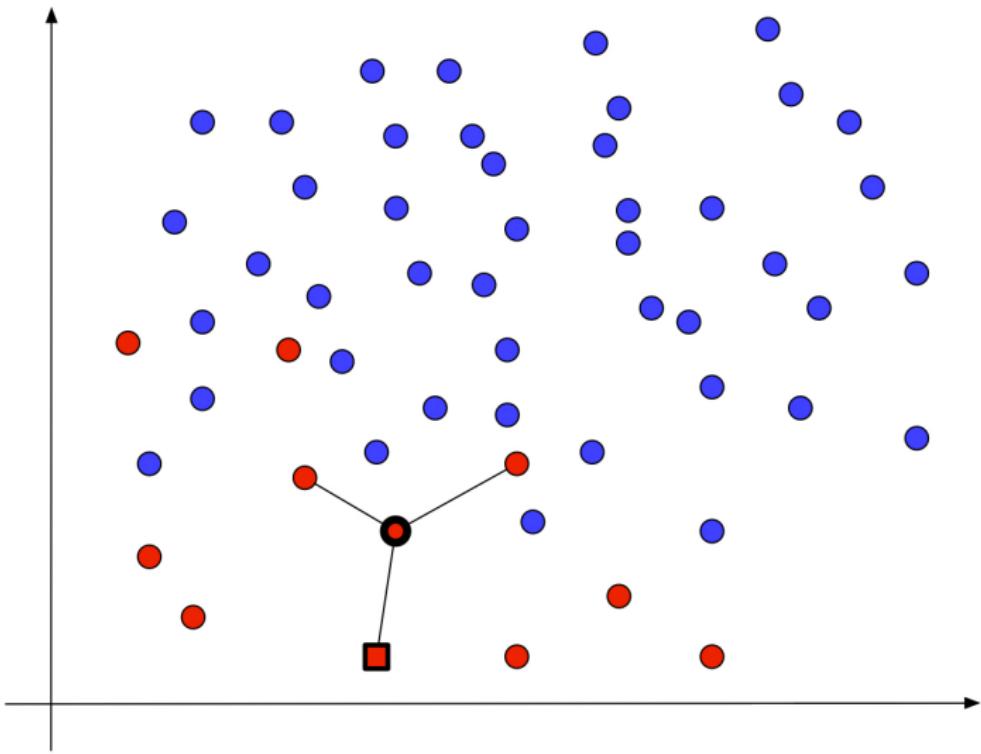
# SMOTE V



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

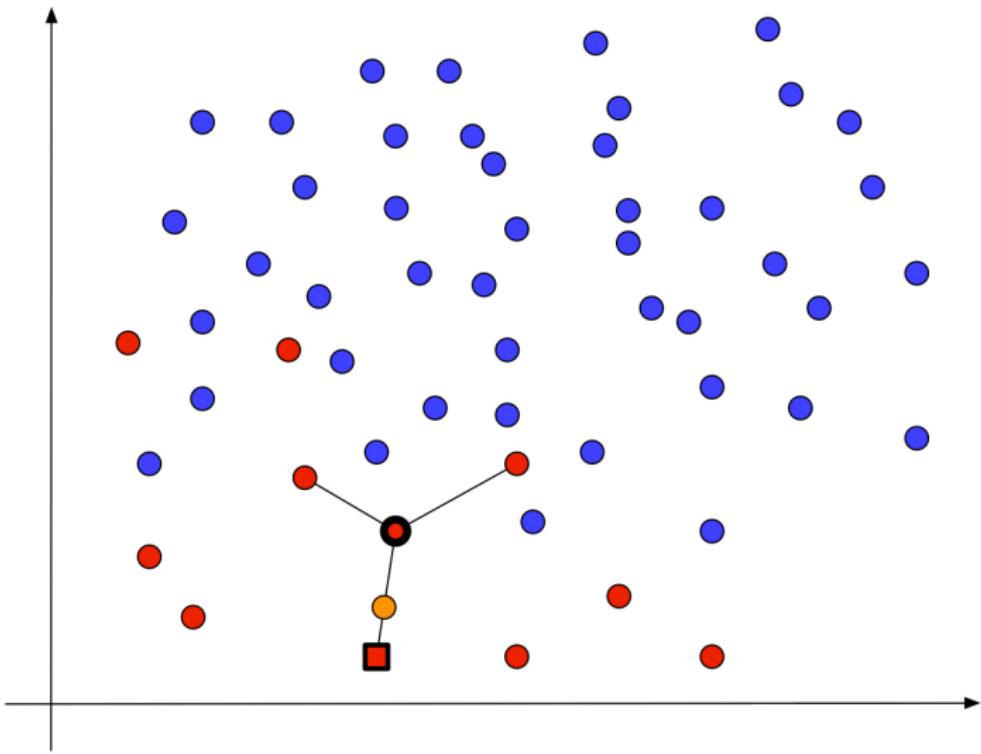
# SMOTE VI



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

# SMOTE VII



- Introduction
- Apprentissage supervisé
- Sur-apprentissage
- Critères d'évaluation de la performance de modèles
- Compléments sur la classification supervisée
- Classification supervisée multi-classes
- Données déséquilibrées

Références

# Sous-échantillonnage

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

On peut procéder à du **sous-échantillonnage** (*under-sampling*) :

- ▶ Méthode naïve : supprimer aléatoirement des observations de la classe majoritaire jusqu'à l'obtention d'un ratio classe minoritaire- classe majoritaire acceptable.
- ▶ Les algorithmes **Tomek link** et **NearMiss** visent à supprimer les observations de la classe majoritaire proches des observations de la classe minoritaire.

## Autres alternatives

- ▶ On peut utiliser le **cost-sensitive learning**, comme le gradient boosting, qui désigne les méthodes d'apprentissage prenant en compte le coût d'une mauvaise classification. Il est ainsi possible d'attribuer des poids accrus pour les observations issues de la classe minoritaire.
- ▶ On peut également utiliser les techniques d'**ensemble learning** (méthodes d'agrégation) : constituer plusieurs échantillons équilibrés à partir des mêmes données de la classe minoritaire et agréger les résultats obtenus.

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Classification supervisée multi-classes

Données déséquilibrées

Références

# Références

Introduction

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Compléments sur la classification supervisée

Références

- Chawla, N. V., K. W. Bowyer, L. O. Hall et W. P. Kegelmeyer. 2002, «Smote : Synthetic minority over-sampling technique», *Journal of Artificial Intelligence Research*, vol. 16, p. 321—357.
- Hastie, T., R. Tibshirani et J. H. Friedman. 2009, *The elements of statistical learning. Data Mining, inference, and prediction*, 2<sup>e</sup> éd., Springer Series in Statistics, Springer.
- James, G., D. Witten, T. Hastie et R. Tibshirani. 2015, *An introduction to statistical learning with applications in R*, Springer Texts in Statistics, Springer.