

Apprentissage statistique

Vincent Lefieux

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références



Plan

Généralités

Généralités

Apprentissage supervisé

Apprentissage
supervisé

Sur-apprentissage

Sur-
apprentissage

Critères d'évaluation de la performance de modèles

Critères
d'évaluation de
la performance
de modèles

Références

Plan

Généralités

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Apprentissage supervisé ou non ?

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

► 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.

Apprentissage supervisé ou non II

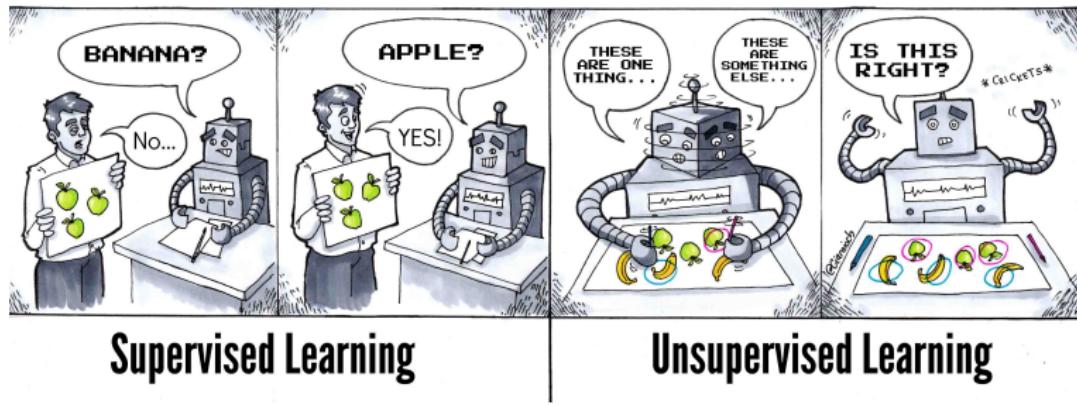
Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références



The difference between supervised and unsupervised learning,
illustrated by [@Ciaraioch](#)

Apprentissage supervisé I

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$...			
$i = 2$...			
$i = 3$...			
.
.
$i = n - 1$...			
$i = n$...			

Apprentissage supervisé II

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$...			
$i = 2$...			
$i = 3$...			
.
$i = n - 1$...			
$i = n$...			

Apprentissage supervisé III

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$...			
$i = 2$...			
$i = 3$...			
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$i = n - 1$...			
$i = n$...			

$$\Rightarrow \forall i \in \{1, \dots, n\} : y_i \approx \hat{f}(x_{i1}, \dots, x_{ip})$$

Apprentissage supervisé IV

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$	■	■	■	...	■	■	■
$i = 2$	■	■	■	...	■	■	■
$i = 3$	■	■	■	...	■	■	■
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$i = n - 1$	■	■	■	...	■	■	■
$i = n$	■	■	■	...	■	■	■
<hr/>							
$i = n + 1$	■	■	■	...	■	■	

$$\Rightarrow \forall i \in \{1, \dots, n\} : y_i \approx \hat{f}(x_{i1}, \dots, x_{ip})$$

Apprentissage supervisé V

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$	■	■	■	...	■	■	■
$i = 2$	■	■	■	...	■	■	■
$i = 3$	■	■	■	...	■	■	■
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
$i = n - 1$	■	■	■	...	■	■	■
$i = n$	■	■	■	...	■	■	■
<hr/>							
$i = n + 1$	■	■	■	...	■	■	■ $\leftarrow y_{n+1}^{(p)} = \hat{f}(x_{n+1,1}, \dots, x_{n+1,p})$

Apprentissage supervisé VI

Généralités

Apprentissage supervisé

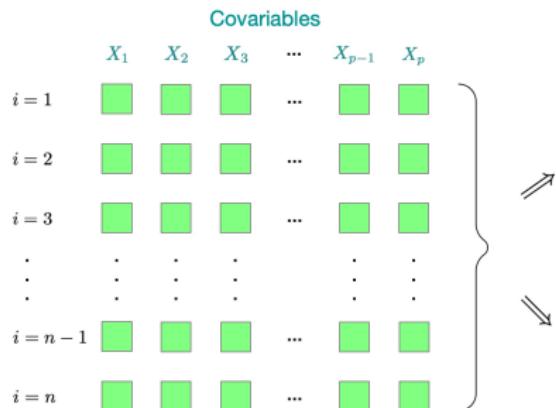
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

	Covariables						Label
	X_1	X_2	X_3	...	X_{p-1}	X_p	Y
$i = 1$	■	■	■	...	■	■	■
$i = 2$	■	■	■	...	■	■	■
$i = 3$	■	■	■	...	■	■	■
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
$i = n - 1$	■	■	■	...	■	■	■
$i = n$	■	■	■	...	■	■	■
<hr/>							
$i = n + 1$	■	■	■	...	■	■	■ $\leftarrow y_{n+1}^{(p)} = \hat{f}(x_{n+1,1}, \dots, x_{n+1,p})$
\vdots	\vdots	\vdots	\vdots		\vdots	\vdots	\vdots

Apprentissage non-supervisé



Analyse factorielle

Relations entre covariables
Réduction de dimension

Classification non supervisée (clustering)

Regroupement d'individus

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

D'autres types d'apprentissage

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

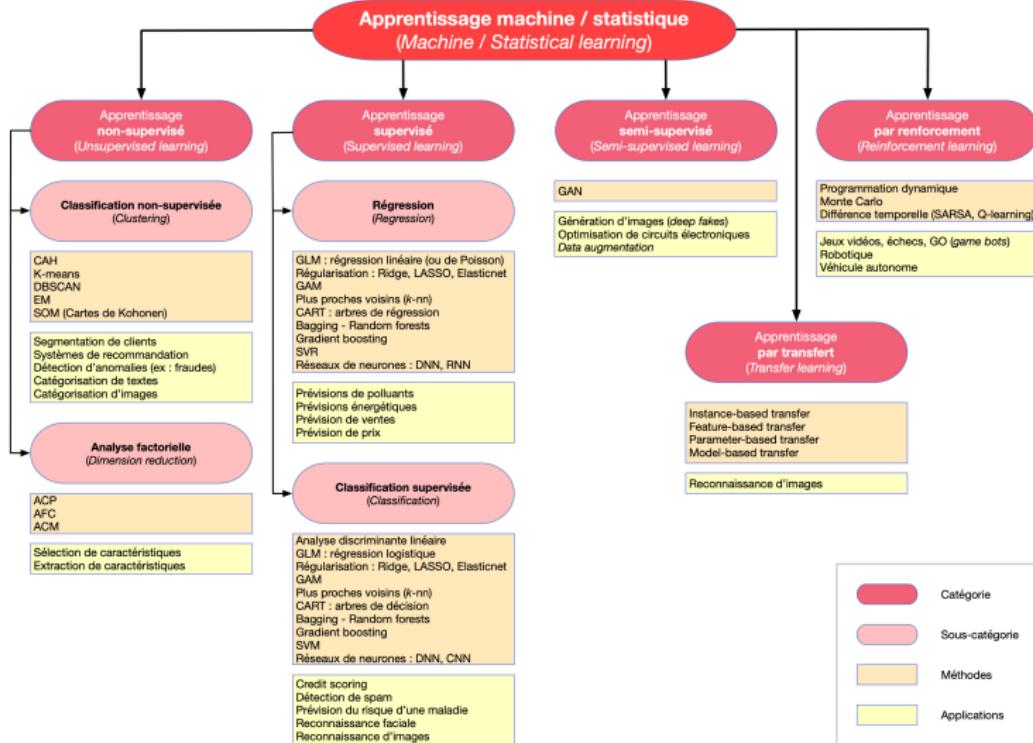
- ▶ Apprentissage **par renforcement** :

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 (avec du **fine tuning**).

Une cartographie



Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

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

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

- ▶ **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.

Modélisation et/ou prévision

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

- ▶ 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.

Quelques enjeux en prévision

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

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**.

Le fléau de la dimension I

Généralités

Apprentissage supervisé

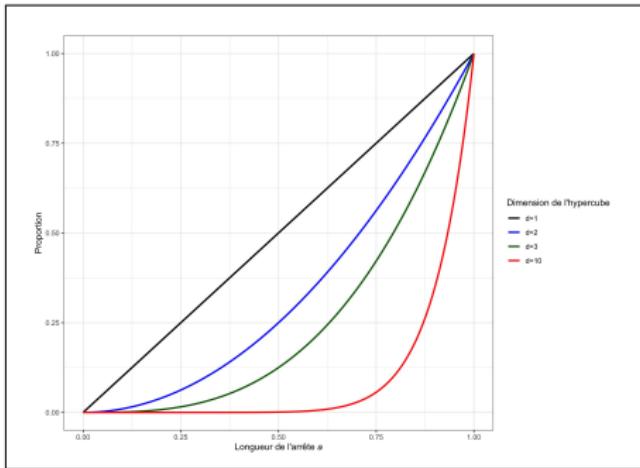
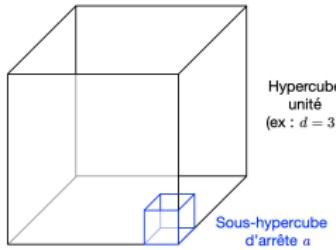
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

- ▶ En grande dimension, il est nécessaire de disposer de volumes de données importants pour « couvrir » les différents axes.
- ▶ Sans données en quantité suffisante, toutes les méthodes basées sur l'estimation locale, comme les arbres et les méthodes d'agrégation, seront incapables d'estimer correctement la fonction de lien (classification supervisée et régression).

Le fléau de la dimension II



- ▶ Un sous-hypercube d'arrête égale à $\frac{1}{10}$ représente $\frac{1}{10^d}$ du volume total en dimension 1, $\frac{1}{100}$ en dimension 2...
- ▶ Si on souhaitait disposer de 10 données à minima sur un sous-hypercube d'arrête $\frac{1}{10}$ en dimension 10, en supposant une répartition uniforme, il faudrait disposer d'un échantillon de taille 10×10^{10} , soit 100 milliards de données.

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Plan

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

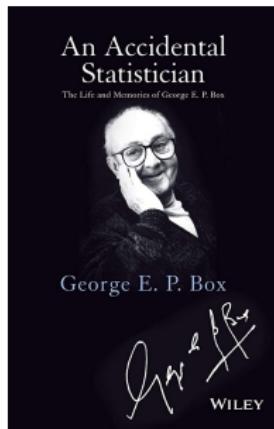
Critères d'évaluation de la performance de modèles

Références

En guise de préambule

All models are wrong,
but some are useful

George E. P. Box



Less is more
Mies van der Rohe



Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Plan

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Données

- ▶ On dispose d'un échantillon de (X_1, \dots, X_p, Y) :

$$d_n = (x_{i1}, \dots, x_{ip}, y_i)_{i \in \{1, \dots, n\}}.$$

- ▶ X_1, \dots, X_p sont appelées **covariables, caractéristiques (features), variables (explicatives)** ou prédicteurs.
- ▶ Y est appelé **variable à expliquer, label ou étiquette**.

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

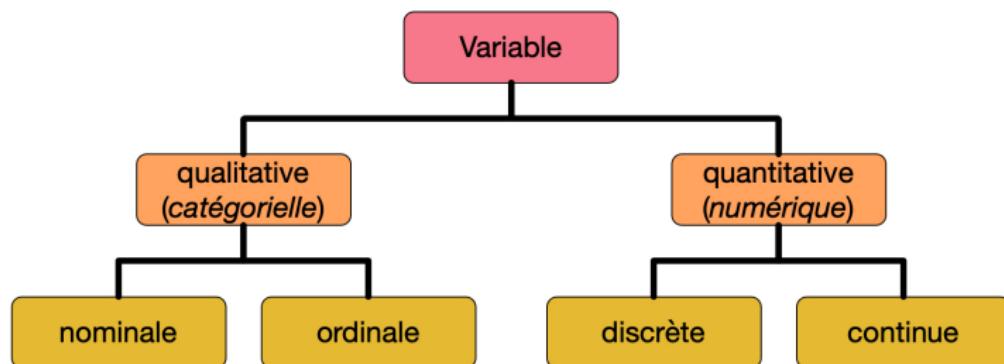
Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Rappel sur les types de variables



Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Objectif

- ▶ On se place dans le cadre de la **prévision** : on souhaite prévoir y pour une nouvelle valeur x .
- ▶ Illustration sur le jeu de données MNIST (<http://yann.lecun.com/exdb/mnist/>) :

- ▶ Modéliser :

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

- ▶ Prévoir :

Covariables

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

- ▶ On considèrera très souvent dans la suite que :

$$X_1, \dots, X_p \in \mathbb{R}^p .$$

- ▶ 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

- ▶ **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 $Y \in \{-1, 1\}$ (ou
 $Y \in \{0, 1\}$).
Par défaut la classification supervisée sera considérée binaire mais on sera amené à traiter des classifications supervisées avec (strictement) plus de 2 labels.

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Généralités

Apprentissage supervisé

Formalisation

Perthes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

- ▶ On suppose que (X, Y) où $X = (X_1, \dots, X_p)$ est 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} .

Plan

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Qualité d'un estimateur

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

- ▶ La qualité d'un estimateur \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**).

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** :

$$\ell(y, y') = \mathbb{1}_{y \neq y'}.$$

- ▶ Cas de la **régression** :

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

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

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Risque (erreur de généralisation)

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

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

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Oracle

- ▶ Si on connaissait $P_{X,Y}$, on pourrait déterminer la fonction de lien optimale, appelé **oracle** :

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

- ▶ Quelques exemples :

- ▶ 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 :

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

Généralités

Apprentissage supervisé

Formalisation

Perthes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Enjeu

- ▶ L'objectif 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 :

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

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Risque empirique

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

- ▶ 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 $R(\hat{f})$.

Plan

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

**Erreurs
d'approximation et
d'estimation**

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Hypothèses de travail et erreurs induites

- ▶ 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 :

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

- ▶ On définit l'**erreur d'approximation** et l'**erreur d'estimation** comme suit :

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

Généralités

Apprentissage supervisé

Formalisation

Perthes et risques

Erreurs d'approximation et d'estimation

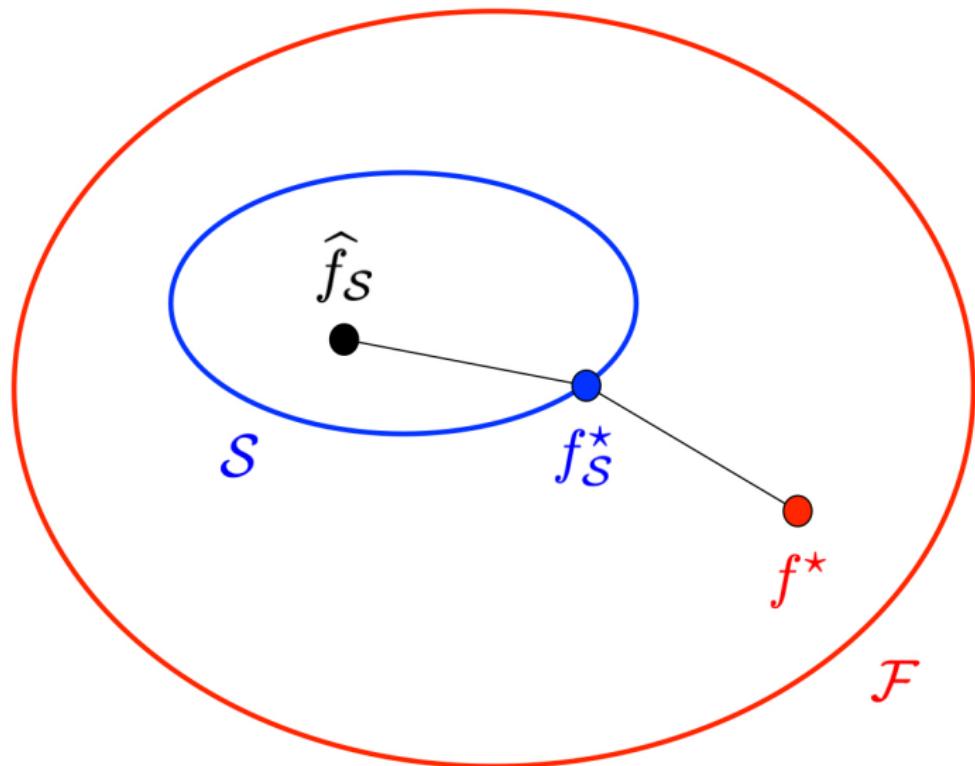
Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Illustration



Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Plan

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Décomposition de l'erreur quadratique I

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

- ▶ On considère le modèle de régression :

$$Y = f(X) + \varepsilon$$

et on estime f par \hat{f} .

- ▶ La **variance** de l'estimateur \hat{f} est définie par :

$$\text{Var} [\hat{f}(X)] = \mathbb{E} \left[(\hat{f}(X) - \mathbb{E} [\hat{f}(X)])^2 \right].$$

- ▶ Le **biais** de l'estimateur \hat{f} est défini par :

$$\text{Biais} (\hat{f}(X)) = \mathbb{E} [\hat{f}(X)] - f(X).$$

Décomposition de l'erreur quadratique II

- ▶ L'**erreur quadratique** de l'estimateur \hat{f} se décompose comme suit :

$$\begin{aligned}& \mathbb{E} \left[(Y - \hat{f}(X))^2 \right] \\&= \mathbb{E} \left[(\hat{f}(X) - \mathbb{E} [\hat{f}(X)])^2 \right] + \left(\mathbb{E} [\hat{f}(X)] - f(X) \right)^2 + \sigma^2 \\&= \text{Var} [\hat{f}(X)] + \text{Biais}^2 (\hat{f}(X)) + \sigma^2\end{aligned}$$

où σ^2 est l'erreur irréductible.

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références

Illustration I

Généralités

Apprentissage
supervisé

Formalisation

Pertes et risques

Erreurs
d'approximation et
d'estimation

Biais et variance

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Prévision de
 $y \in \mathcal{Y}$

Echantillon 1

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

• • •

Echantillon K

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

Illustration II

Généralités

Apprentissage supervisé

Formalisation

Pertes et risques

Erreurs d'approximation et d'estimation

Biais et variance

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Références



En orange : $\hat{y}^{(p)1}, \dots, \hat{y}^{(p)K}$

Plan

Sur-apprentissage

Exemple

Problème

Palliatifs

Généralités

Apprentissage
supervisé

**Sur-
apprentissage**

Exemple

Problème

Palliatifs

Critères
d'évaluation de
la performance
de modèles

Références

Plan

Sur-apprentissage

Exemple

Problème

Palliatifs

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Exemple

Problème

Palliatifs

Critères
d'évaluation de
la performance
de modèles

Références

Impact de la complexité : exemple I

- ▶ On considère le modèle de régression suivante :

$$\forall x \in [0, 1] : y = \cos\left(\frac{3\pi x}{2}\right) + \varepsilon$$

où $\varepsilon \sim \mathcal{N}(0, 0.1^2)$.

- ▶ On simule :

- ▶ $n_{train} = 30$ observations $(x_i, y_i)_{i=1, \dots, n_{train}}$ à partir desquelles on estime cette fonction de régression par régression linéaire polynomiale de degré $d \in \{1, \dots, 11\}$; on obtient les valeurs ajustées $(\hat{y}_i)_{i=1, \dots, n_{train}}$.
- ▶ $n_{test} = 8$ observations $(x_i, y_i)_{i=1, \dots, n_{test}}$, non utilisées pour l'apprentissage (l'estimation) du modèle, pour lesquelles on obtient les prévisions $(\hat{y}_i^{(p)})_{i=1, \dots, n_{test}}$ qu'on compare aux réalisations $(y_i)_{i=1, \dots, n_{test}}$.

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple II

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

- ▶ On évalue la qualité :

- ▶ de l'ajustement par le RMSE (*Root Mean Squared Error*) sur l'échantillon d'apprentissage :

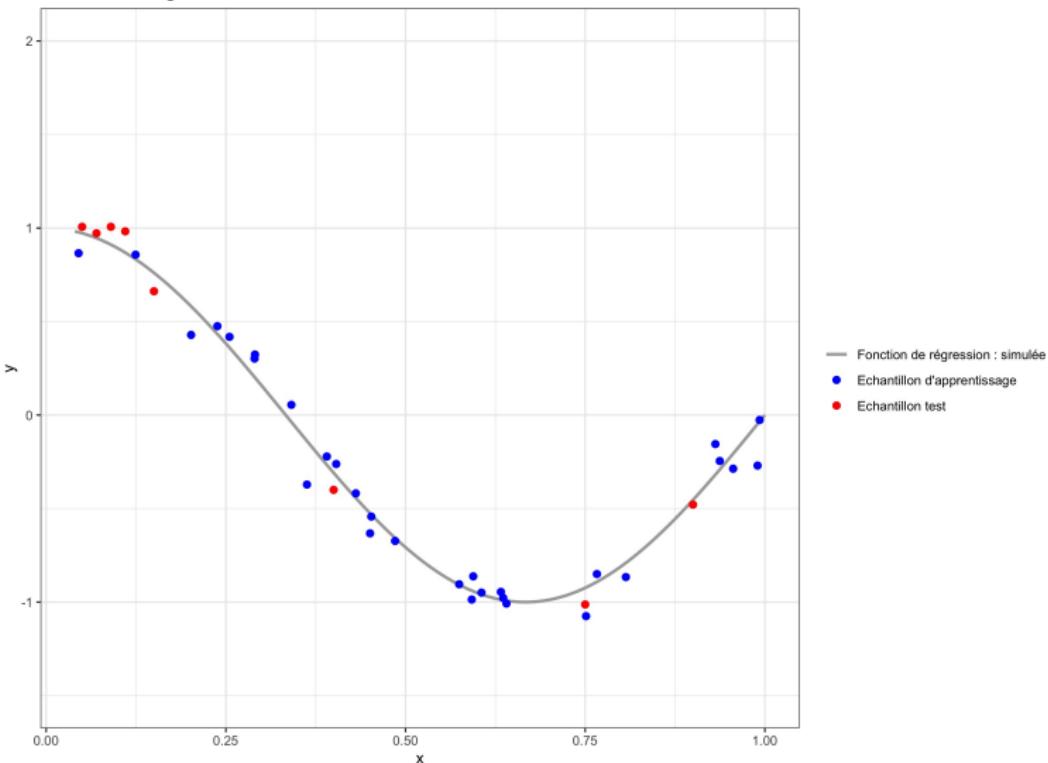
$$\text{RMSE}_{train} = \sqrt{\frac{1}{n_{train}} \sum_{i=1}^{n_{train}} (\hat{y}_i - y_i)^2}.$$

- ▶ de la prévision par le RMSE sur l'échantillon de test :

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

Impact de la complexité : exemple III

Fonction de régression & échantillons



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

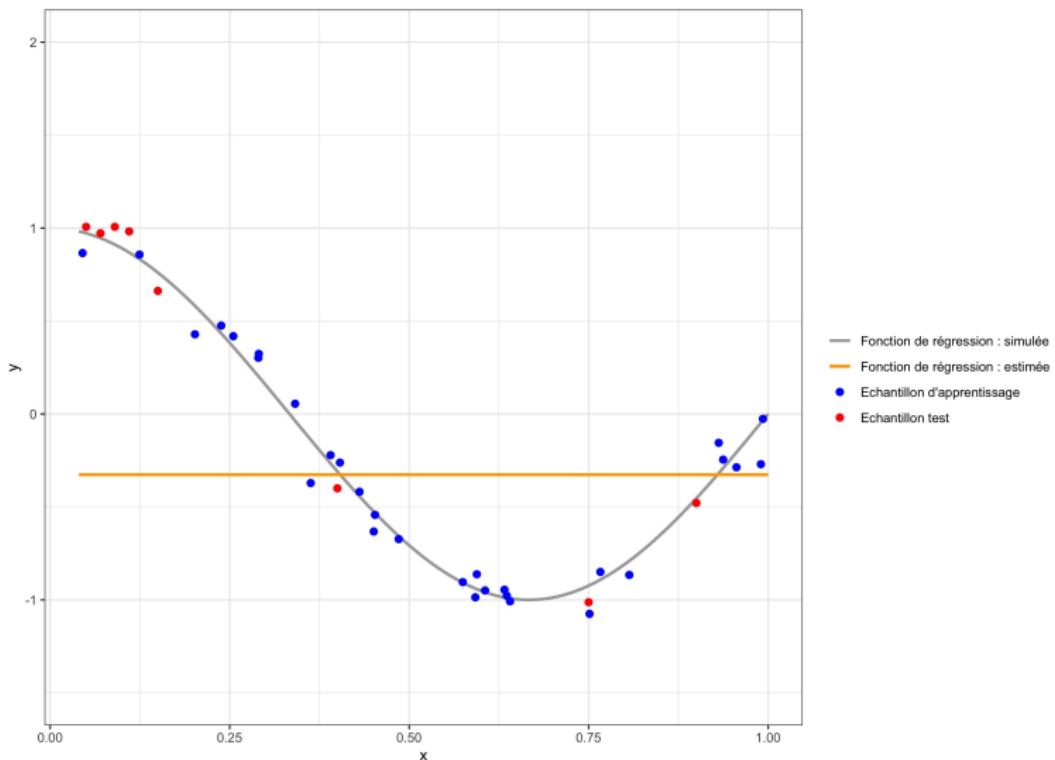
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple IV

Régression polynômale de degré 0 : RMSE_{train} = 0.57 ; RMSE_{test} = 1.03



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

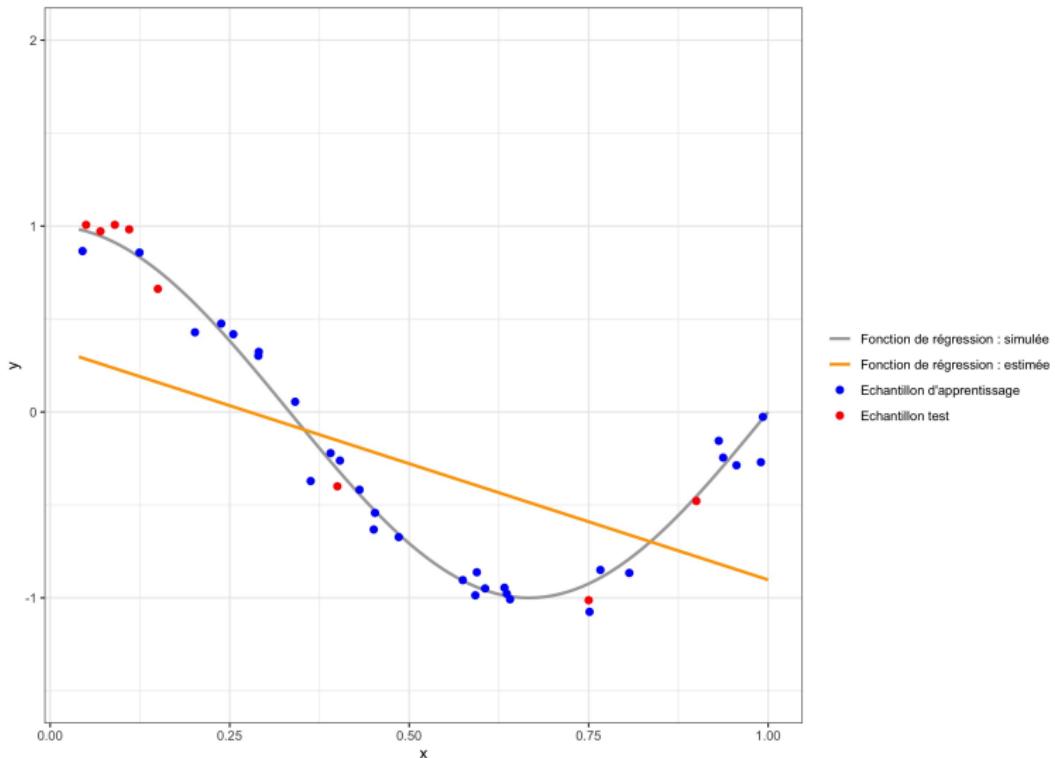
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple V

Régression polynômale de degré 1 : RMSE_{train} = 0.47 ; RMSE_{test} = 0.59



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

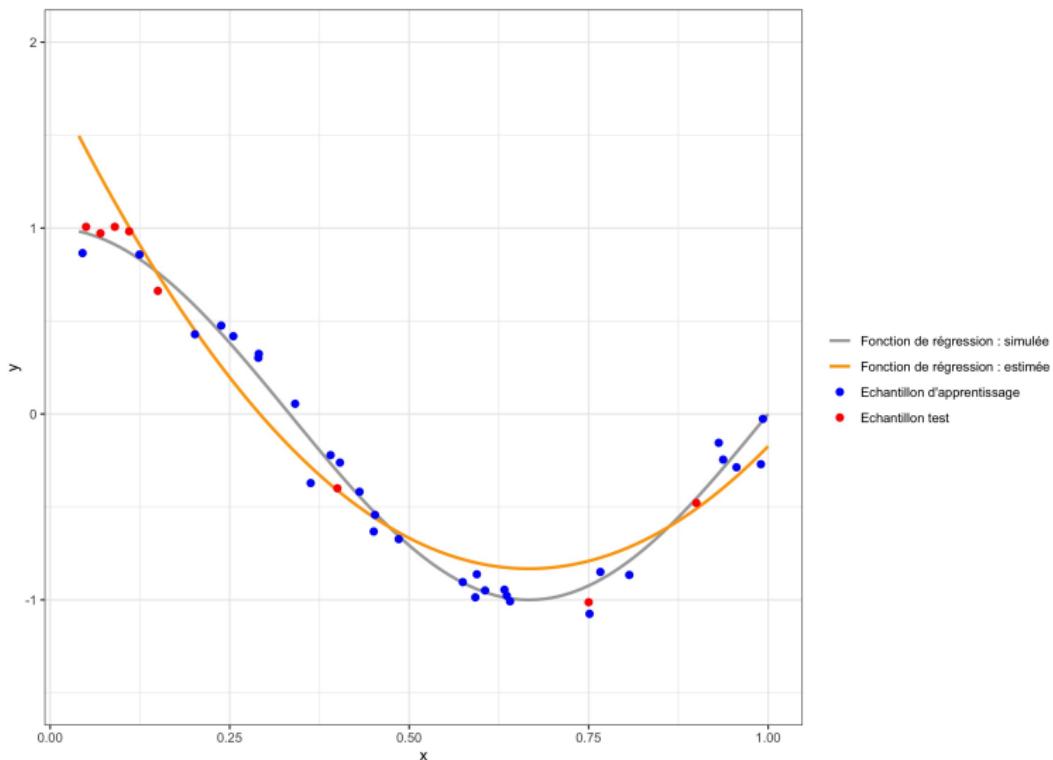
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple VI

Régression polynômiale de degré 2 : $\text{RMSE}_{\text{train}} = 0.2$; $\text{RMSE}_{\text{test}} = 0.21$



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

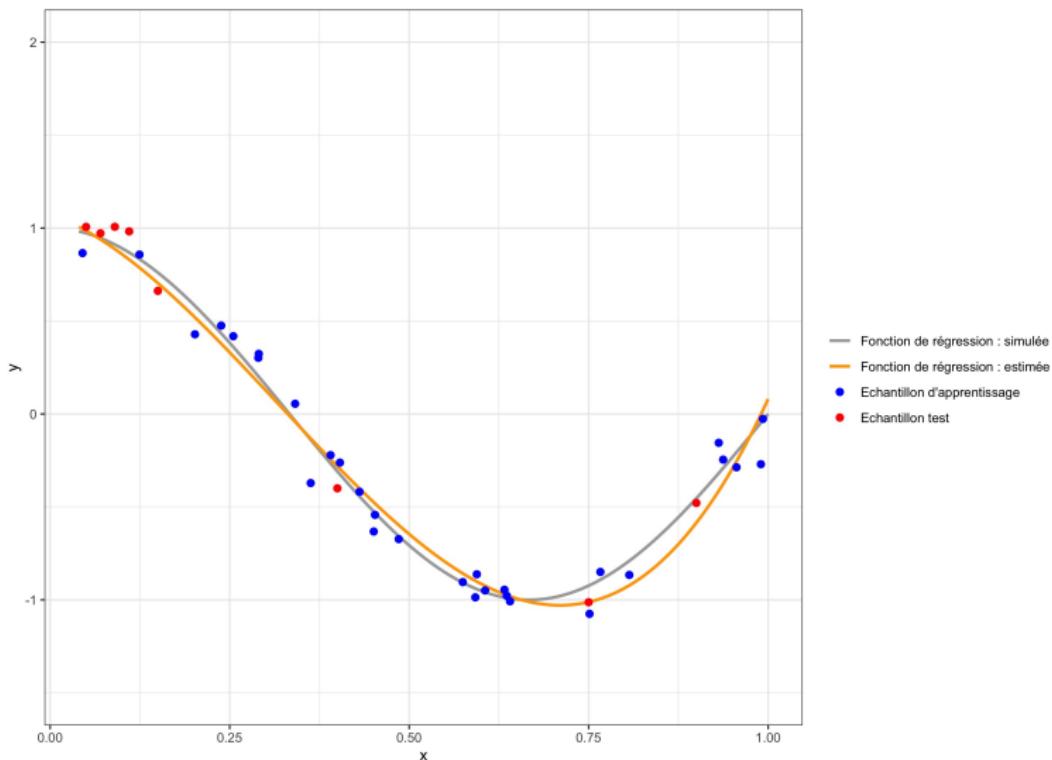
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple VII

Régression polynômale de degré 3 : $\text{RMSE}_{\text{train}} = 0.12$; $\text{RMSE}_{\text{test}} = 0.09$



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

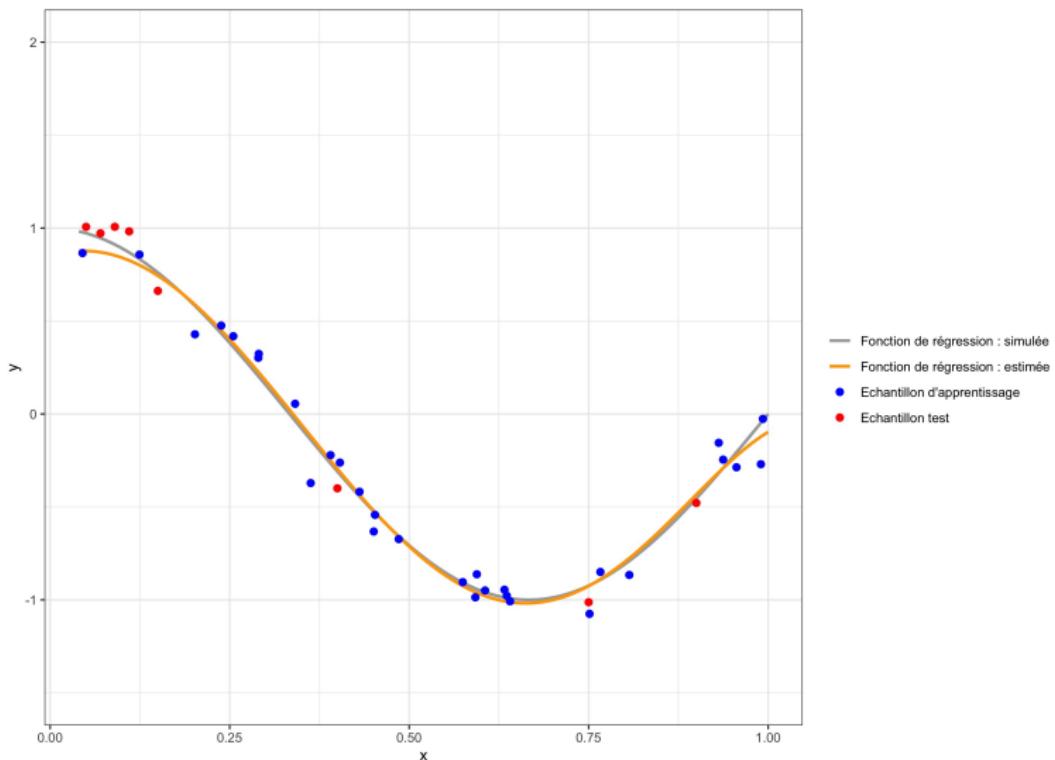
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple VIII

Régression polynômale de degré 5 : $\text{RMSE}_{\text{train}} = 0.09$; $\text{RMSE}_{\text{test}} = 0.11$



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

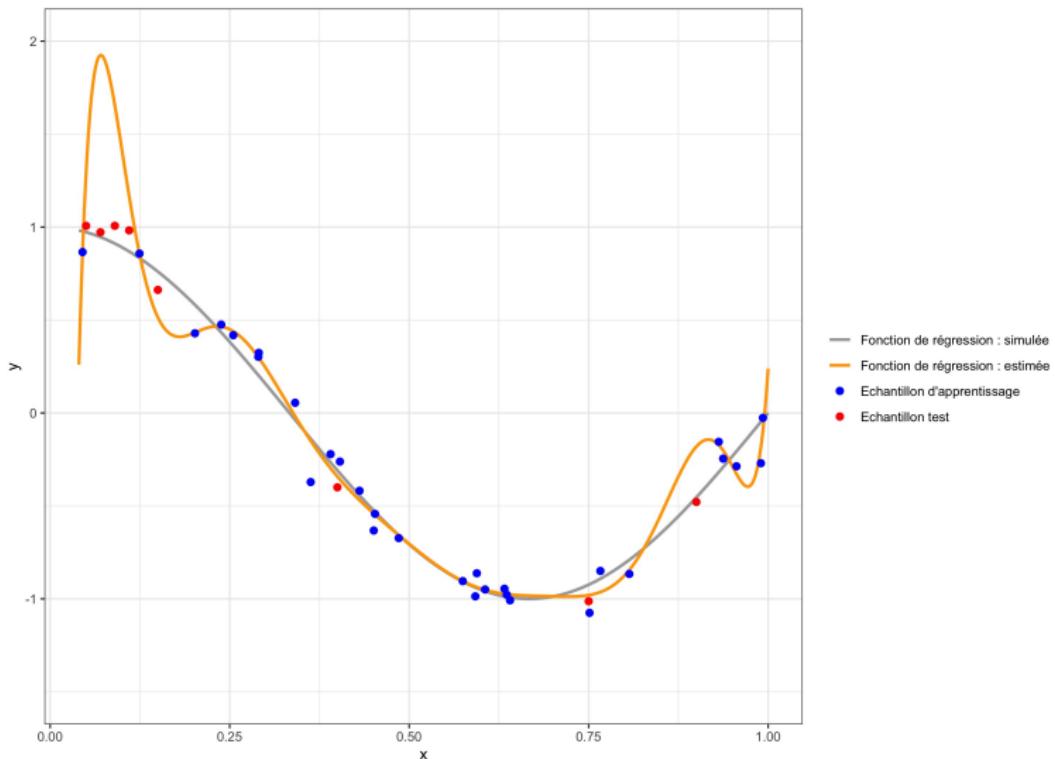
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple IX

Régression polynômiale de degré 11 : $\text{RMSE}_{\text{train}} \approx 0.07$; $\text{RMSE}_{\text{test}} \approx 0.44$



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

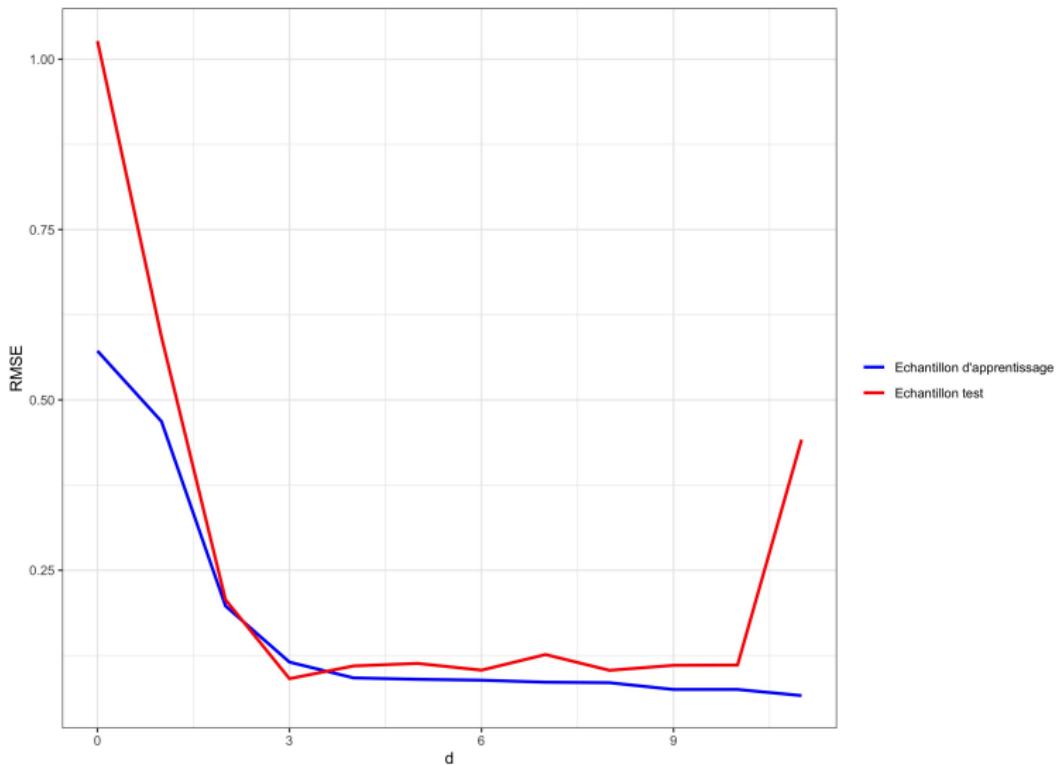
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Impact de la complexité : exemple X



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Plan

Sur-apprentissage

Exemple

Problème

Palliatifs

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Exemple

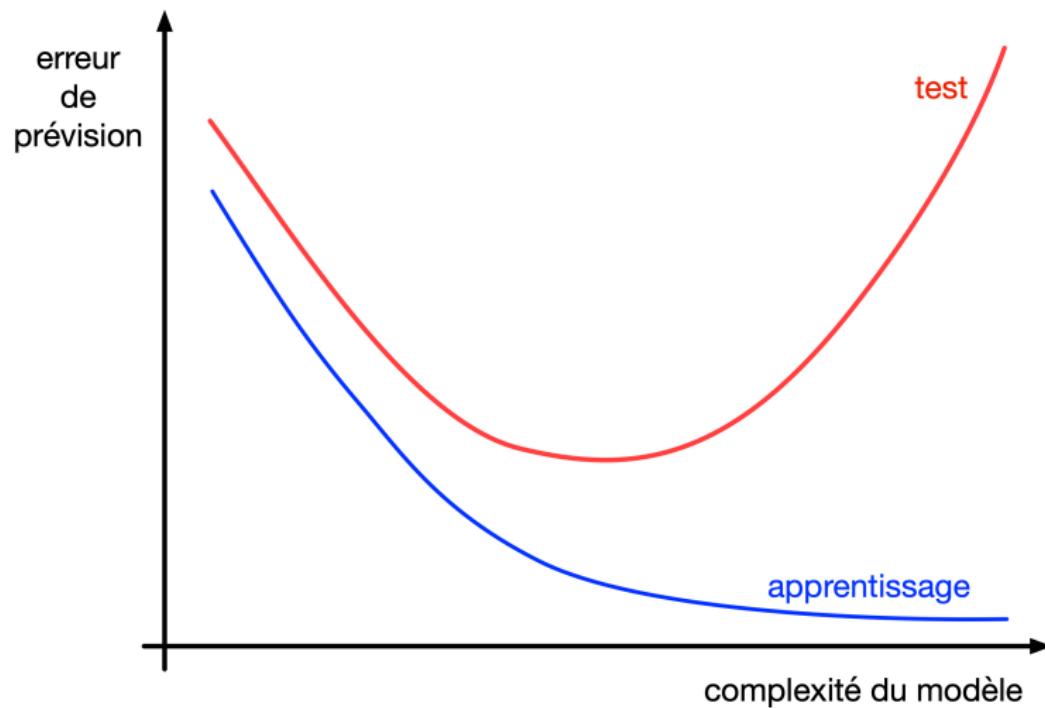
Problème

Palliatifs

Critères
d'évaluation de
la performance
de modèles

Références

Erreur de prévision et complexité I



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

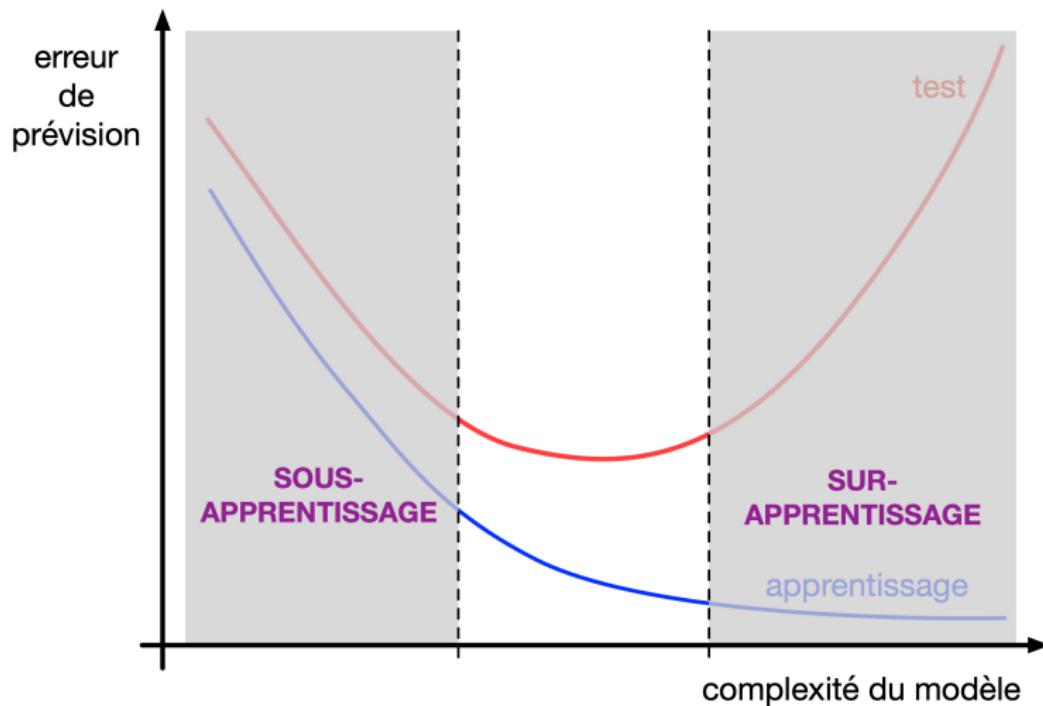
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Erreur de prévision et complexité II



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Caractérisation de la complexité

- ▶ Pour le **modèle linéaire** (régression linéaire ou régression logistique) : le **nombre de covariables**.
- ▶ Pour les **méthodes de régularisation** (Ridge, LASSO, Elastic Net) : **inverse(s) du(des) paramètre(s) de régularisation**.
- ▶ Pour **CART** : le **nombre de découpages**.
- ▶ Pour les **random forests** : aucun, il n'y a pas de sur-apprentissage (le temps de calcul reste néanmoins une limite).
- ▶ Pour le **gradient boosting** : le **nombre d'itérations**.
- ▶ Dans les **réseaux de neurones** : la **taille du réseau** et le **nombre d'époques**.

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

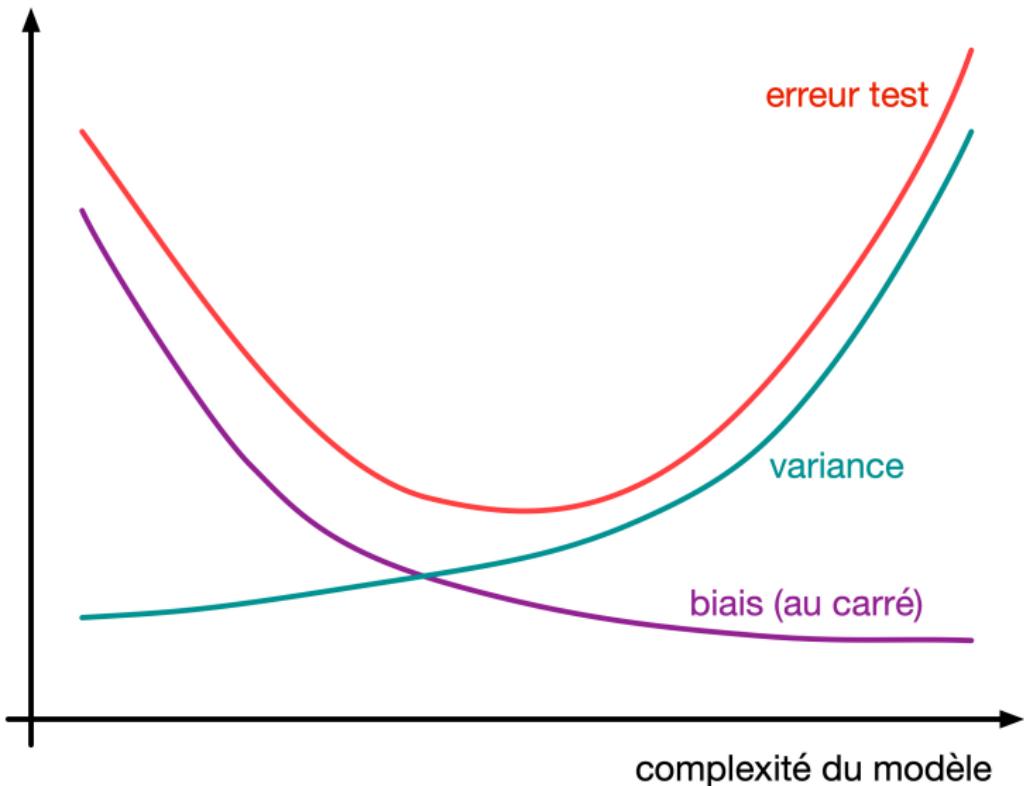
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Erreur test, biais et variance



Généralités
Apprentissage supervisé
Sur-apprentissage
Exemple
Problème
Palliatifs
Critères d'évaluation de la performance de modèles
Références

Plan

Sur-apprentissage

Exemple

Problème

Palliatifs

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Exemple

Problème

Palliatifs

Critères
d'évaluation de
la performance
de modèles

Références

Plusieurs stratégies

- ▶ Utilisation de **critères d'information** tels que l'AIC et le BIC (ou le C_p de Mallows) pour les **modèles linéaires** (régression linéaire et régression logistique).
- ▶ Méthodes de rééchantillonage :
 - ▶ Les procédures sont :
 - ▶ **Apprentissage / Validation / Test** : dans le cas où on dispose de **gros échantillons**.
 - ▶ **Validation croisée** : dans le cas où on dispose d'**échantillons de taille moyenne**.
 - ▶ **Validation croisée itérée** : dans le cas où on dispose de **petits échantillons**.
 - ▶ L'échantillon de **validation** permet de calibrer les **hyperparamètres optimaux**.
 - ▶ L'échantillon **test** permet d'évaluer l'**erreur de prévision** à partir des données disponibles.

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

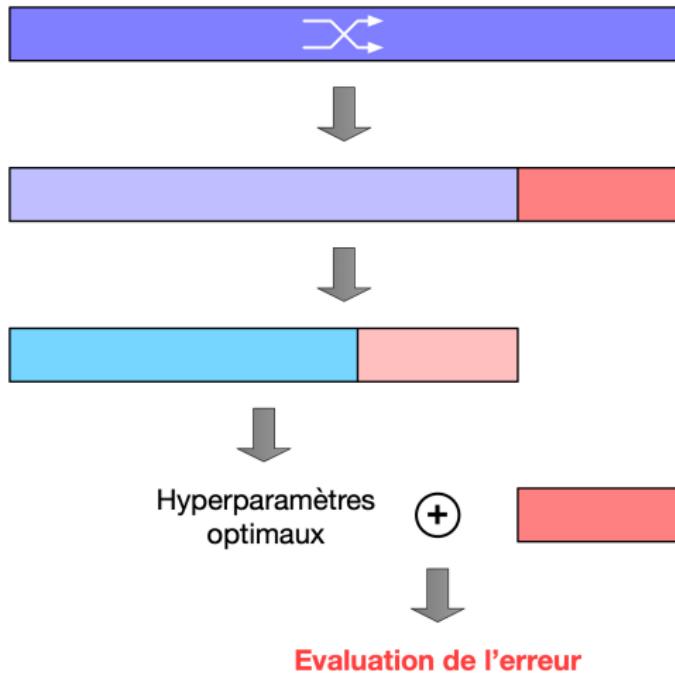
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

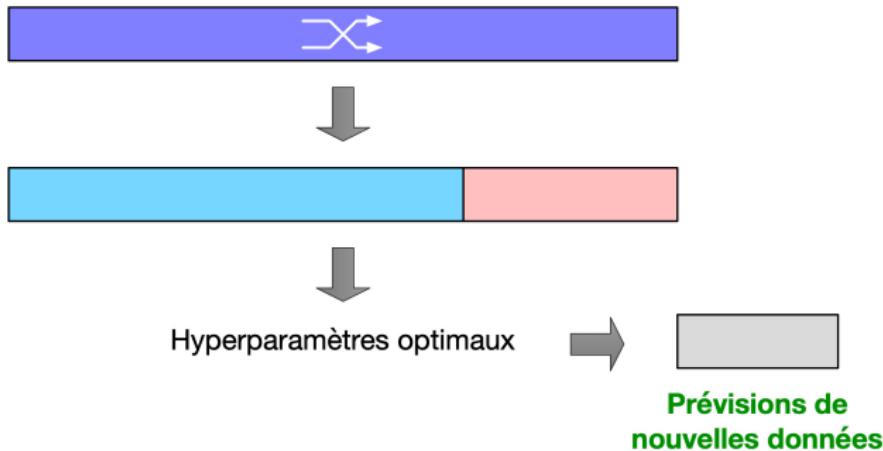
Références

Apprentissage/validation/test : erreur



Généralités
Apprentissage supervisé
Sur-apprentissage
Exemple
Problème
Palliatifs
Critères d'évaluation de la performance de modèles
Références

Apprentissage/validation/test : prévision



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée I

- 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 .

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

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

- Calculer le critère de validation croisée :

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

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée II

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

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

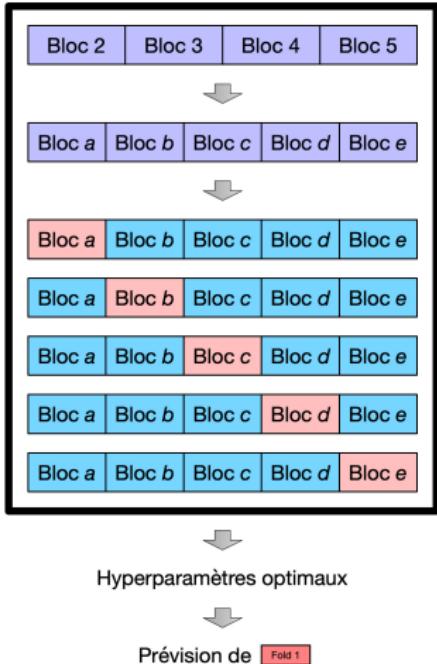
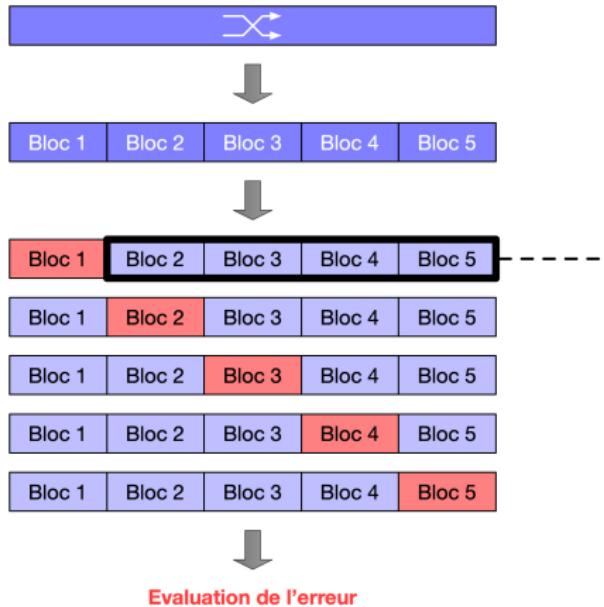
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) : erreur



Tout Apprentissage et validation Test

Apprentissage Validation

Généralités

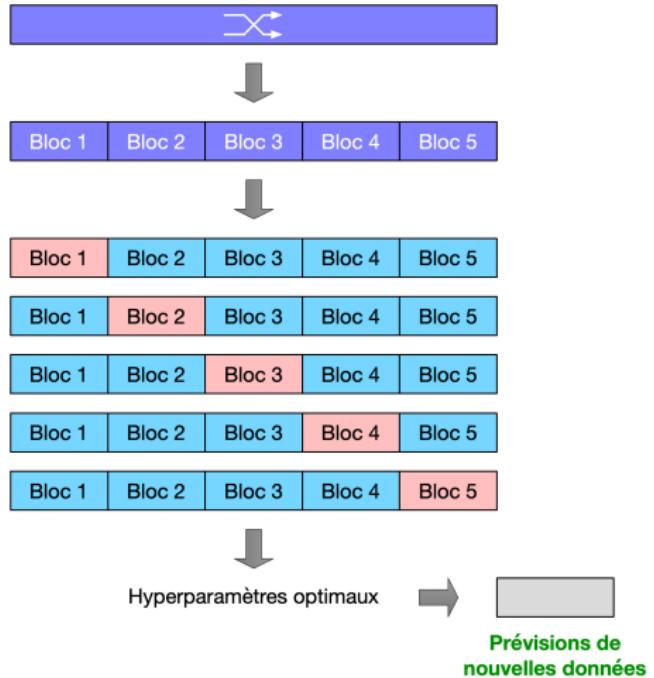
Apprentissage supervisé

Sur-apprentissage
Exemple
Problème
Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) : prévision



Tout Apprentissage Validation

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

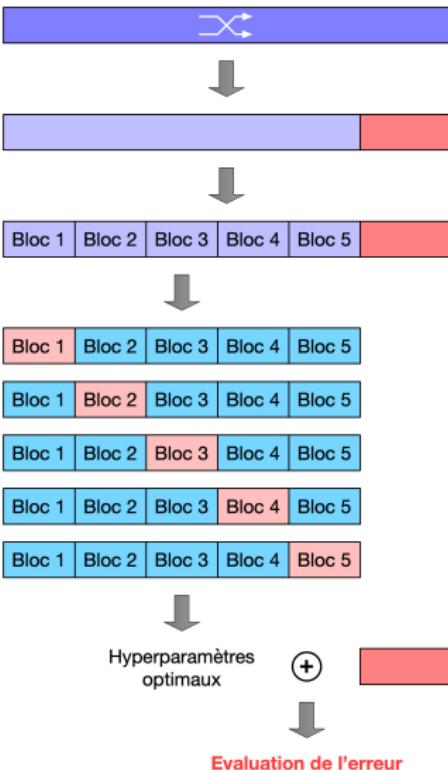
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) partielle : erreur



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

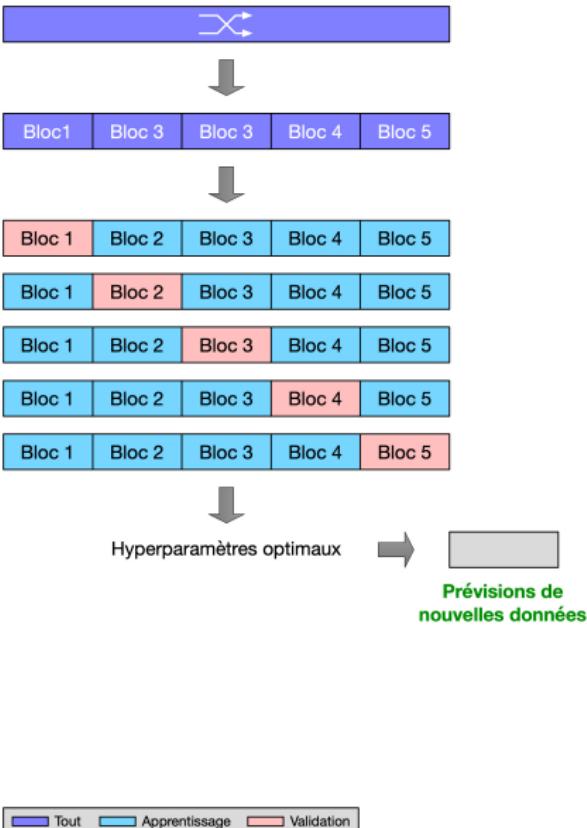
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) partielle : prévision



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

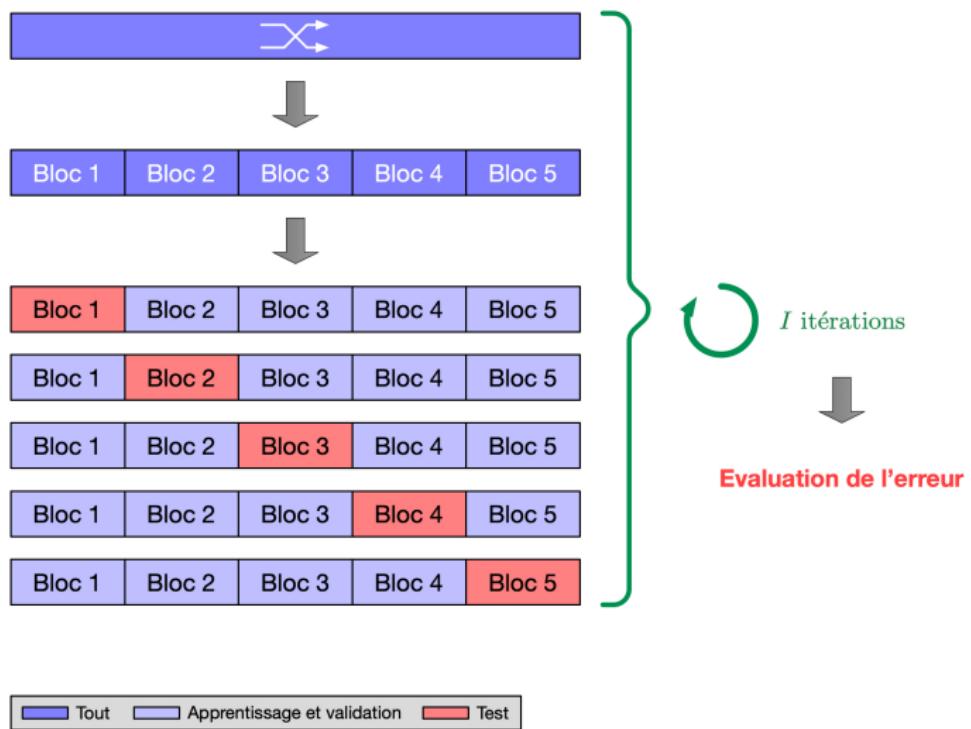
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) itérée : erreur



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

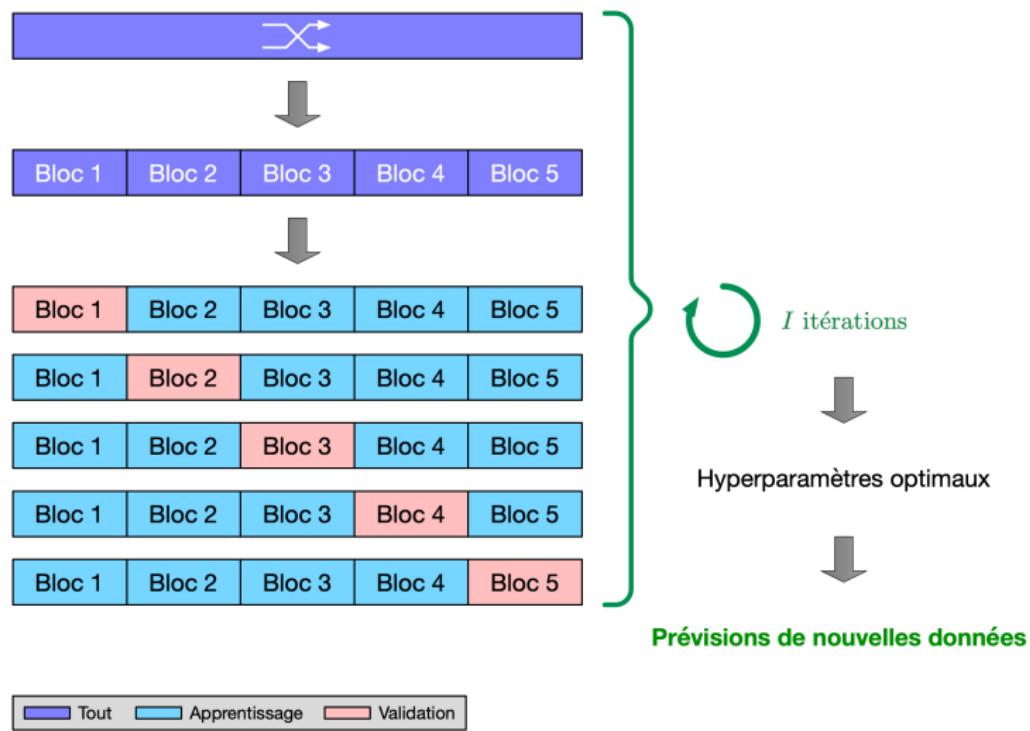
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Validation croisée ($K = 5$) itérée : prévision



Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

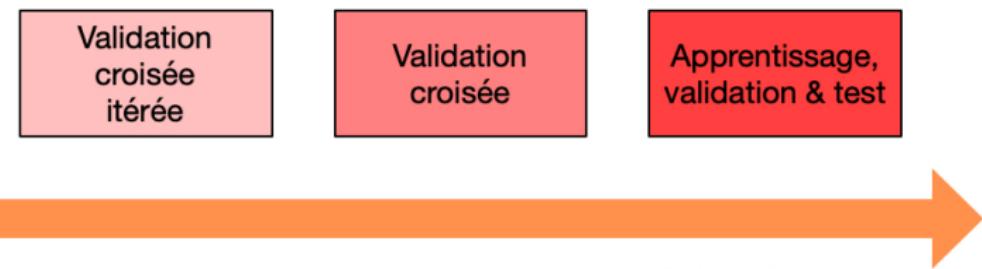
Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Choix de la stratégie



Taille de l'échantillon
Coût calculatoire de l'algorithme

Généralités

Apprentissage supervisé

Sur-apprentissage

Exemple

Problème

Palliatifs

Critères d'évaluation de la performance de modèles

Références

Plan

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Plan

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Prévision (cas de la classification supervisée binaire)

- ▶ On procède de la manière suivante :
 - ▶ On **estime** un modèle sur l'**échantillon d'apprentissage**.
 - ▶ On utilise ce modèle estimé pour calculer les **probabilités prévues** $\widehat{\mathbb{P}}(Y = 1 / X = x_i)$ sur l'**échantillon test**.
 - ▶ On en déduit les **prévisions** $\widehat{y}_i^{(p)}$ sur l'**échantillon test** :

$$\widehat{y}_i^{(p)} = \begin{cases} 1 & \text{si } \widehat{\mathbb{P}}(Y = 1 / X = x_i) \geq s \\ 0 \text{ (ou -1)} & \text{sinon} \end{cases}$$

où s est un seuil de décision fixé par l'utilisateur (usuellement $s = \frac{1}{2}$).

- ▶ S'il existe de nombreux critères de qualité possibles, il faut choisir celui(ceux) qui répond(ent) au mieux à la problématique métier.

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Matrice de confusion

- Dans le cas de la classification supervisée binaire, la **matrice de confusion** vaut (avec les notations TP et TN pour True Positive et True negative) :

		Prévision	
		1	0 (ou -1)
Vérité	1	Vrai Positif (TP)	Faux Négatif (FN)
	0 (ou -1)	Faux Positif (FP)	Vrai Négatif (TN)

- Dans le cas de la classification supervisée non-binaire, on peut établir la matrice de confusion, avec autant de lignes et de colonnes que de classes, et en déduire les nombres TP, FP, TN et FN.

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Critères de qualité

- ▶ 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.

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Classification
supervisée

Régression

Références

Exactitude (et erreur de classification) et spécificité

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

$$\text{exactitude} = \frac{TP + TN}{TP + FP + TN + FN}.$$

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

$$\text{erreur} = \frac{FP + FN}{TP + FP + TN + FN} = 1 - \text{exactitude}.$$

- ▶ La **spécificité** (*specificity*), le taux de négatifs classés négatifs (« taux de vrais négatifs » noté **TNR** : True Negative Rate), vaut :

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

L'**anti-spécificité** est le taux de négatifs classés positifs (« taux de faux positifs » noté **FPR** : False Positive Rate).

Précision, sensibilité et F_1

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

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

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

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

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

$$\begin{aligned}\mathbf{F}_1 &= \frac{2}{\frac{1}{\text{précision}} + \frac{1}{\text{sensibilité}}} \\ &= 2 \frac{\text{précision} \cdot \text{sensibilité}}{\text{précision} + \text{sensibilité}}.\end{aligned}$$

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Courbe ROC I

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

$$\hat{y}_i^{(p)} = \begin{cases} 1 & \text{si } \hat{\mathbb{P}}(Y = 1 / X = x_i) \geq s \\ 0 \text{ (ou -1)} & \text{sinon} \end{cases}.$$

- ▶ 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 $s \in]0, 1]$.

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Courbe ROC II

Généralités

Apprentissage supervisé

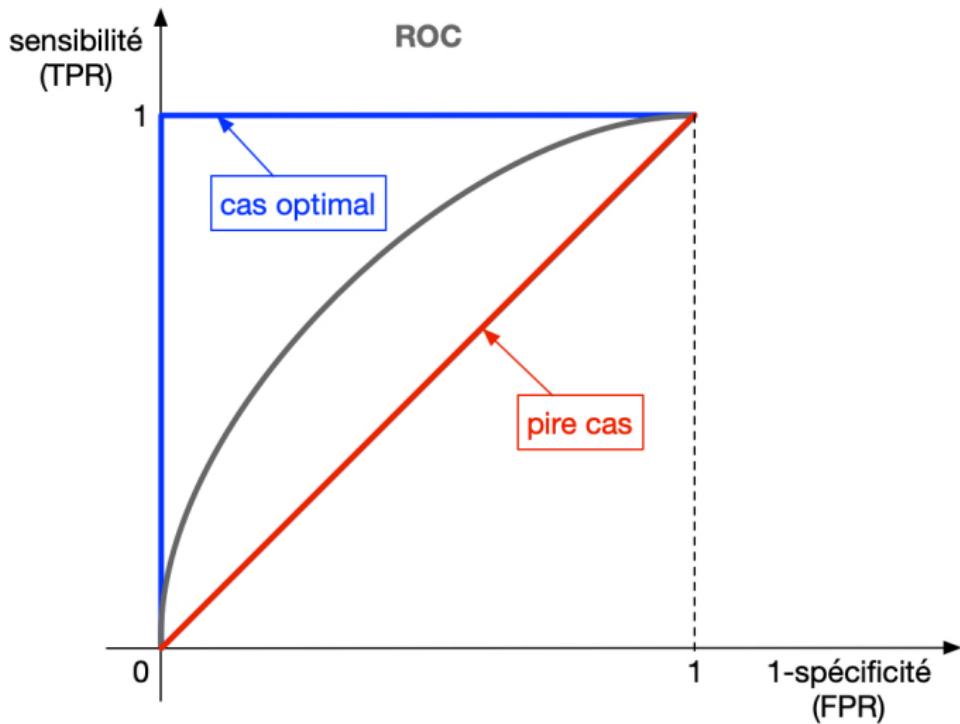
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références



AUC I

Généralités

Apprentissage supervisé

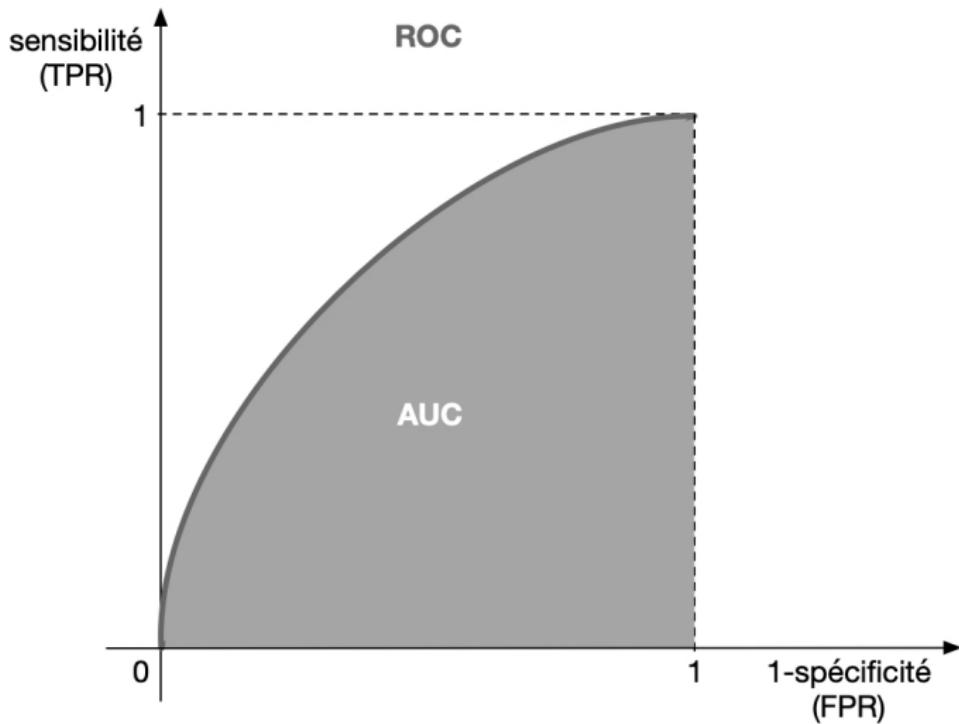
Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références



AUC II

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

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 »).

Exemple ROC/AUC I

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Individus	Label	Probabilité d'être positif
1	0	0.10
2	0	0.15
3	1	0.20
4	0	0.30
5	0	0.35
6	1	0.35
7	0	0.40
8	1	0.55
9	1	0.60
10	1	0.75
11	1	0.75
12	0	0.80
13	1	0.90
14	1	0.95

Exemple ROC/AUC II

Individus	Label	Probabilité d'être positif	Label prévu (avec $s = 0.25$)
1	0	0.10	0
2	0	0.15	0
3	1	0.20	0
4	0	0.30	1
5	0	0.35	1
6	1	0.35	1
7	0	0.40	1
8	1	0.55	1
9	1	0.60	1
10	1	0.75	1
11	1	0.75	1
12	0	0.80	1
13	1	0.90	1
14	1	0.95	1

$$P = 8, \quad N = 6,$$

$$FP = 4, \quad FPR = \frac{FP}{N} = \frac{4}{6},$$

$$TP = 7, \quad TPR = \frac{TP}{P} = \frac{7}{8}.$$

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Exemple ROC/AUC III

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

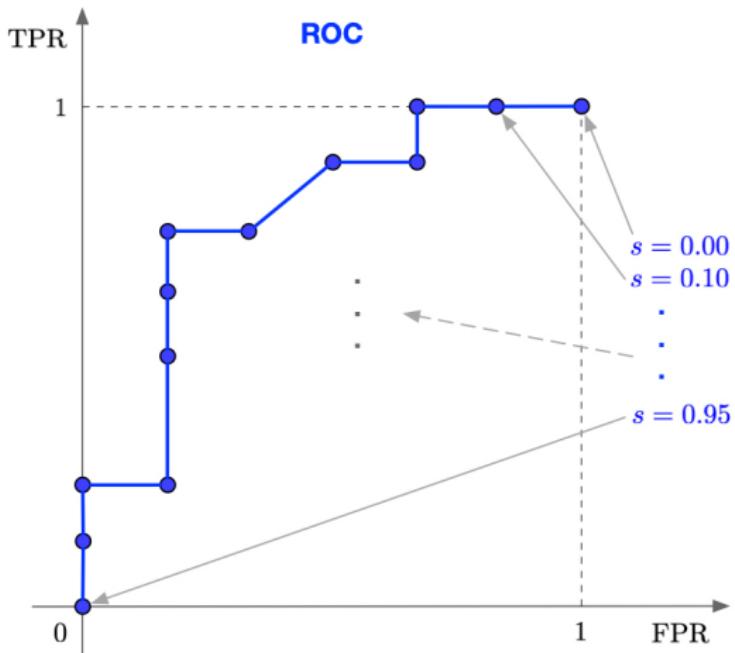
Classification supervisée

Régression

Références

Seuil	Taux de vrais positifs (TP)	Taux de faux positifs (FP)
0.00	$\frac{8}{8}$	$\frac{6}{6}$
0.10	$\frac{8}{8}$	$\frac{5}{6}$
0.15	$\frac{8}{8}$	$\frac{4}{6}$
0.20	$\frac{7}{8}$	$\frac{4}{6}$
0.30	$\frac{7}{8}$	$\frac{3}{6}$
0.35	$\frac{6}{8}$	$\frac{2}{6}$
0.40	$\frac{6}{8}$	$\frac{1}{6}$
0.55	$\frac{5}{8}$	$\frac{1}{6}$
0.60	$\frac{4}{8}$	$\frac{1}{6}$
0.75	$\frac{2}{8}$	$\frac{1}{6}$
0.80	$\frac{2}{8}$	$\frac{0}{6}$
0.90	$\frac{1}{8}$	$\frac{0}{6}$
0.95	$\frac{0}{8}$	$\frac{0}{6}$

Exemple ROC/AUC IV



$$\text{AUC} = \frac{1}{6} \frac{1}{4} + \frac{1}{6} \frac{3}{4} + \frac{1}{2} \frac{1}{6} \left(\frac{6}{8} + \frac{7}{8} \right) + \frac{1}{6} \frac{7}{8} + \frac{1}{3} = \frac{75}{96} \simeq 0.78.$$

Généralités
Apprentissage supervisé
Sur-apprentissage
Critères d'évaluation de la performance de modèles
Classification supervisée
Régression
Références

Plan

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

Prévision

- ▶ On procède de la manière suivante :
 - ▶ On **estime** un modèle sur l'**échantillon d'apprentissage**.
 - ▶ On utilise ce modèle estimé pour calculer les **prévisions** $\hat{y}_i^{(p)}$ sur l'**échantillon test**.
- ▶ Concernant les critères d'erreur :
 - ▶ S'il existe de nombreux critères possibles, il faut choisir celui(ceux) qui répond(ent) au mieux à la problématique métier.
 - ▶ Les **écart quadratiques** sont classiquement utilisés à l'aide des **RMSE** et **nRMSE** (ils correspondent souvent au critère qui a été optimisé).
 - ▶ On peut utiliser des **écart absolu**s à l'aide du **MAE** ou **absolu**s relatifs à l'aide du **MAPE** (à éviter si Y prend des valeurs proches de 0).

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

MSE, RMSE et nRMSE

- Le **MSE** (*Mean Squared Error* : erreur quadratique moyenne) vaut :

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

- Le **RMSE** (*Root Mean Squared Error* : racine carrée de l'erreur quadratique moyenne) vaut :

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

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

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

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

MAE et MAPE

Généralités

Apprentissage supervisé

Sur-apprentissage

Critères d'évaluation de la performance de modèles

Classification supervisée

Régression

Références

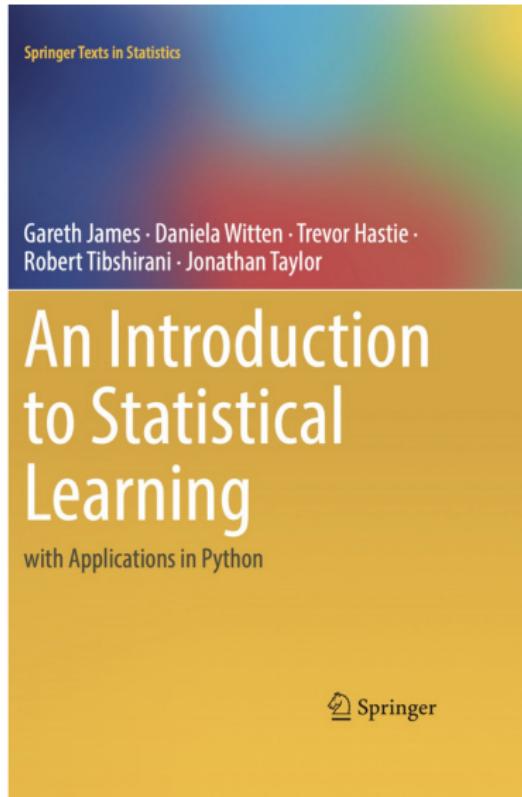
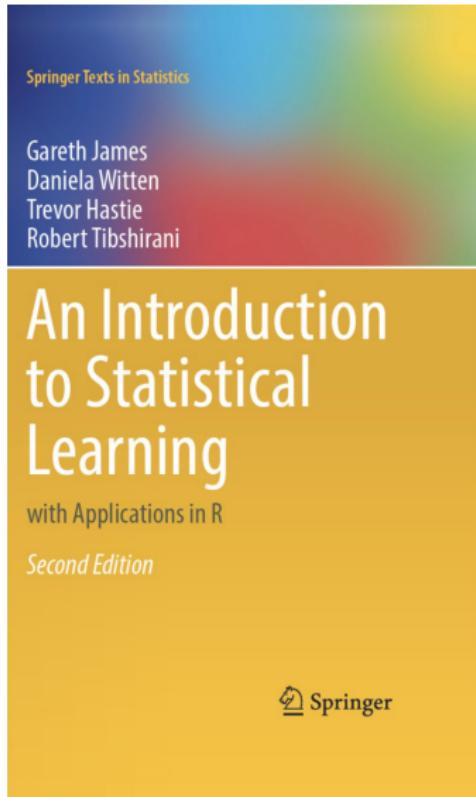
- Le **MAE** (*Mean Absolute Error* : erreur absolue moyenne) vaut :

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

- Le **MAPE** (*Mean Absolute Percent Error* : erreur absolue moyenne relative) vaut :

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

Références I



Généralités

Apprentissage
supervisé

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Classification
supervisée

Régression

Références

Références II

Généralités

Apprentissage
supervisé

Sur-
apprentissage

Critères
d'évaluation de
la performance
de modèles

Références

Hastie, T., R. Tibshirani et J. H. Friedman. 2009, *The elements of statistical learning. Data Mining, inference, and prediction*, 2^e éd., Springer Series in Statistics, Springer.

James, G., D. Witten, T. Hastie et R. Tibshirani. 2021, *An introduction to statistical learning with applications in R*, 2^e éd., Springer Texts in Statistics, Springer.

James, G., D. Witten, T. Hastie, R. Tibshirani et J. Taylor. 2023, *An introduction to statistical learning with applications in Python*, Springer Texts in Statistics, Springer.