

# Métriques de performances

1





# Sommaire

- ▶ **I/ Métriques de performance pour la régression**
- ▶ II/ Métriques de performance pour la classification
- ▶ III/ Techniques de validation



## 1. Notations

- ▶  $N$  - Nombre d'observations
- ▶  $y \in \mathbb{R}^N$  - variable objectif
- ▶  $\hat{y} \in \mathbb{R}^N$  - prédiction
- ▶  $\hat{y}_i \in \mathbb{R}$  - prédiction pour la  $i$ -ème observation
- ▶  $y_i \in \mathbb{R}$  - valeur réel pour la  $i$ -ème observation



## 2. Mean Square Error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

- ▶ La métrique la plus populaire pour mesurer les erreurs de régression
- ▶ L'erreur est difficile à interpréter parce que nous mesurons le carré de l'erreur



### 3. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

- ▶ Chaque minimiseur pour MSE est un minimiseur pour RMSE
- ▶ L'échelle de l'erreur est la même que l'échelle de la cible
- ▶ Mais il est un peu plus facile de travailler avec MSE



## 4. R-squared

$$R^2 = 1 - \frac{MSE}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

- ▶ Chaque minimiseur pour RMSE et MSE est un minimiseur pour R-square
- ▶ Avec le RMSE et le MSE, il est difficile d'estimer si notre modèle est assez bon
- ▶ R-carré est compris entre 0 (modèle pauvre) et 1 (modèle parfait)



## 5. Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

- ▶ RMSE, MSE, R-carré pénalisant les grosses erreurs plus
- ▶ La MAE est moins sensible aux valeurs aberrantes que le R-carré, le MSE et le RMSE



## 6. Mean Square Percentage Error

$$MSPE = \frac{100\%}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2$$

Predicted	Sold	MSE	MSPE
9	10	1	1
999	1000	1	0,0001





## 7. Mean Absolute Percentage Error

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Predicted	Sold	MAE	MAPE
9	10	1	10
999	1000	1	0,1



## 8. Résumons

- ▶ MSE, RMSE, R-squared
  - ▶ Ils sont les mêmes du point de vue de l'optimisation
- ▶ MAE
  - ▶ Robuste aux outliers
- ▶ (R)MSPE
  - ▶ Version pondérée du (R)MSE
- ▶ MAPE
  - ▶ Version pondérée du MAE



# Sommaire

- ▶ I/ Métriques de performance pour la régression
- ▶ **II/ Métriques de performance pour la classification**
- ▶ III/ Techniques de validation



## 1. Notations

- ▶  $N$  - number of observations
- ▶  $L$  - number of classes
- ▶  $y$  - ground truth
- ▶  $\hat{y}$  - predictions
- ▶  $[a = b]$  - indicator function
- ▶ 'soft labels' (soft predictions) - classifier's scores
- ▶ 'hard labels' (hard predictions) - classifier's Boolean
  - ▶  $\arg \max_i f_i(x)$
  - ▶  $[f(x) > b]$ ,  $b$ -threshold



## 2. Accuracy score



Accuracy = 0.99 !!

$$Accuracy = \frac{1}{N} \sum_{i=1}^N [\hat{y}_i = y_i]$$

- ▶ A besoin de prediction 1 ou 0 pas de probabilité
- ▶ Fréquence à laquelle notre prédiction de classe est correcte.
- ▶ Précision entre 0 et 1, plus elle est élevée, mieux c'est.

Labels	Prédictions
10 cats	0 cats
990 dogs	1000 dogs



### 3. Cross-entropy loss or log loss

Binary:

$$\text{Logloss} = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$

Multiclass:

$$\text{Logloss} = -\frac{1}{N} \sum_{i=1}^N \sum_{l=1}^L y_{il} \log(\hat{y}_{il})$$



## 4. Matrice de confusion

		Label prédit	
		$\hat{y} = 0$	$\hat{y} = 1$
Vrai label	$y = 0$	6 TN	9 FP
	$y = 1$	1 FN	10 TP

$$\text{Specificity} = \frac{TN}{TN+FP} = 1 - \text{FPR}$$

$$\text{FPR} = \frac{FP}{FP+TN} = 1 - \text{Specificity}$$

$$\text{Sensitivity, Recall} = \frac{TP}{TP+FN} = \text{TPR}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

FP = False Positive

FN = False Negative

TP = True Positive

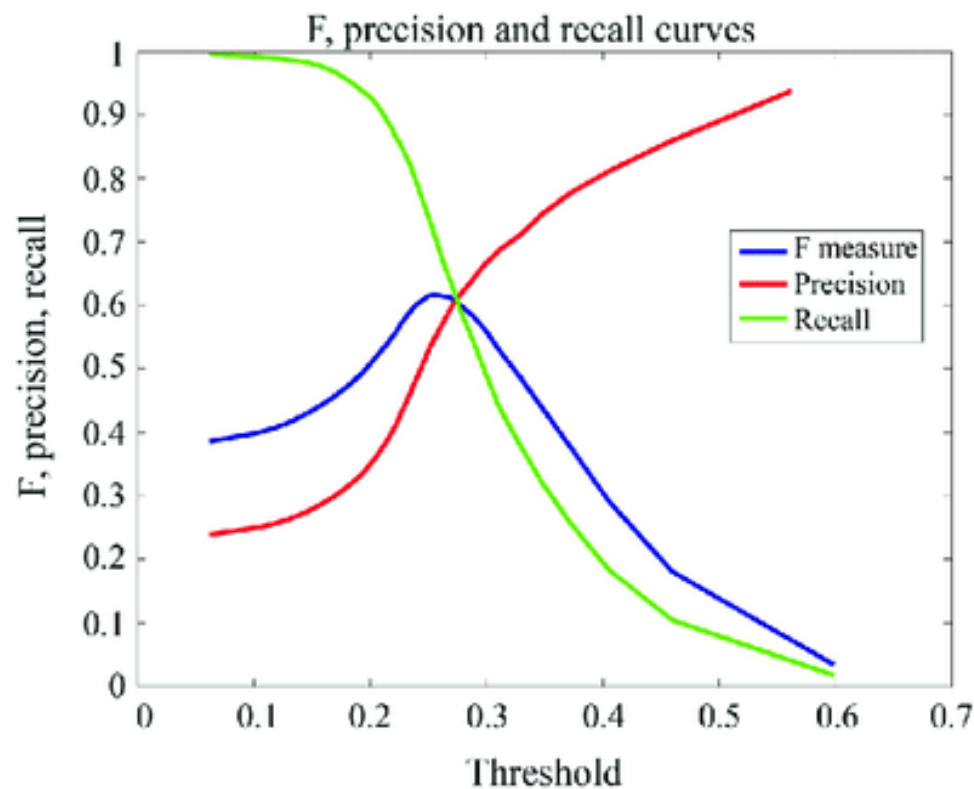
TN = True Negative

FPR = False Positive Rate

TPR = True Positive Rate



## 5. Precision, Recall, F score



$$\text{Recall, Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

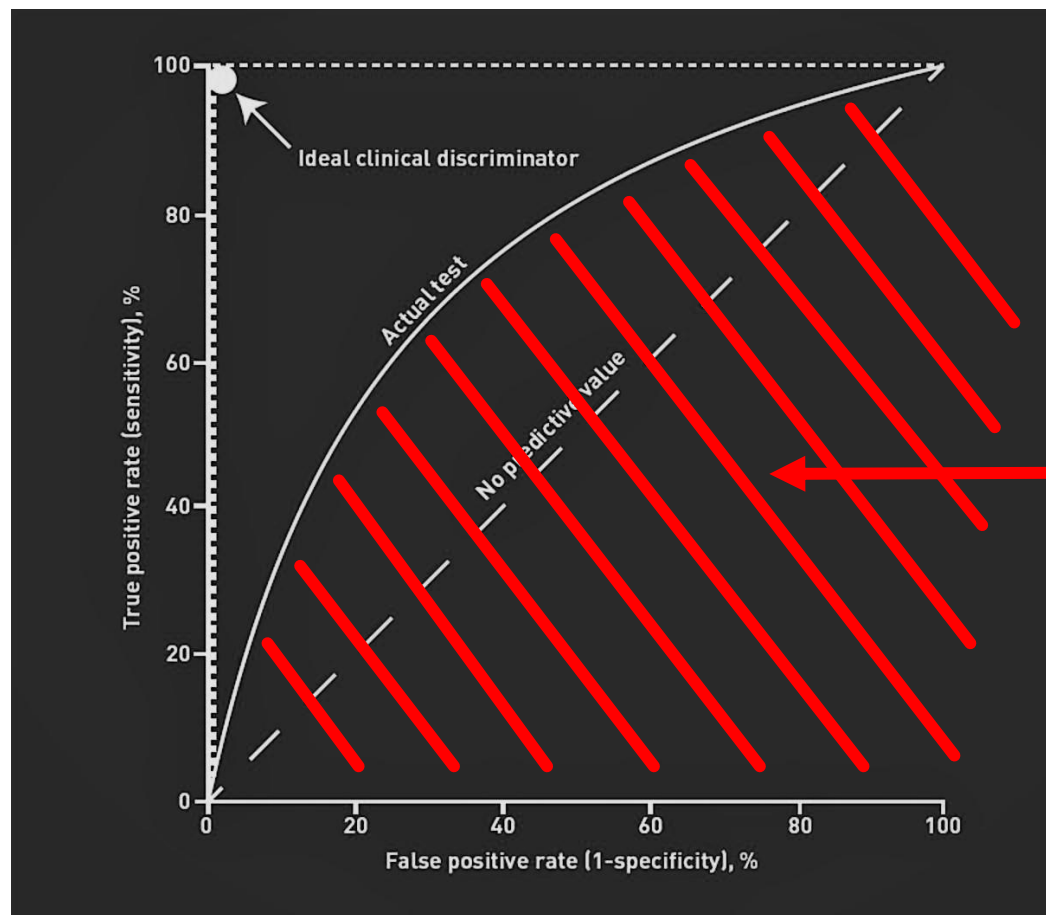
$$F_{\beta} = (1 + \beta) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

$$F_1 = F = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$





## 6. AUC & ROC curve



Area Under Curve (AUC)  
Aire en dessous de la courbe



# Sommaire

- ▶ I/ Métriques de performance pour la régression
- ▶ II/ Métriques de performance pour la classification
- ▶ **III/ Techniques de validation**



# 1. Généralisation

- ▶ Dans le machine learning, l'objectif est de créer un algorithme qui a d'excellentes performances avec de nouvelles données. Nous appelons ce concept le pouvoir de généralisation. Pour mesurer la généralisation de notre modèle, nous allons prédire des données que notre algorithme n'a pas vues au cours de son apprentissage et voir comment il se comporte sur cet ensemble.

## 2. Train and test

		Size ( $x_1$ )	Nb of room ( $x_2$ )	Year ( $x_3$ )	Price ( $y$ )
Jeu d'entraînement (70%)	1	70	3	2010	460
	2	40	3	2015	232
	3	45	4	1990	315
	4	12	2	2017	178
	...	...	...	...	...
Jeu de test (30%)	m-2	60	3	2010	390
	m-1	35	2	1994	300
	m	25	1	2005	240

Training data prediction of the price of a house



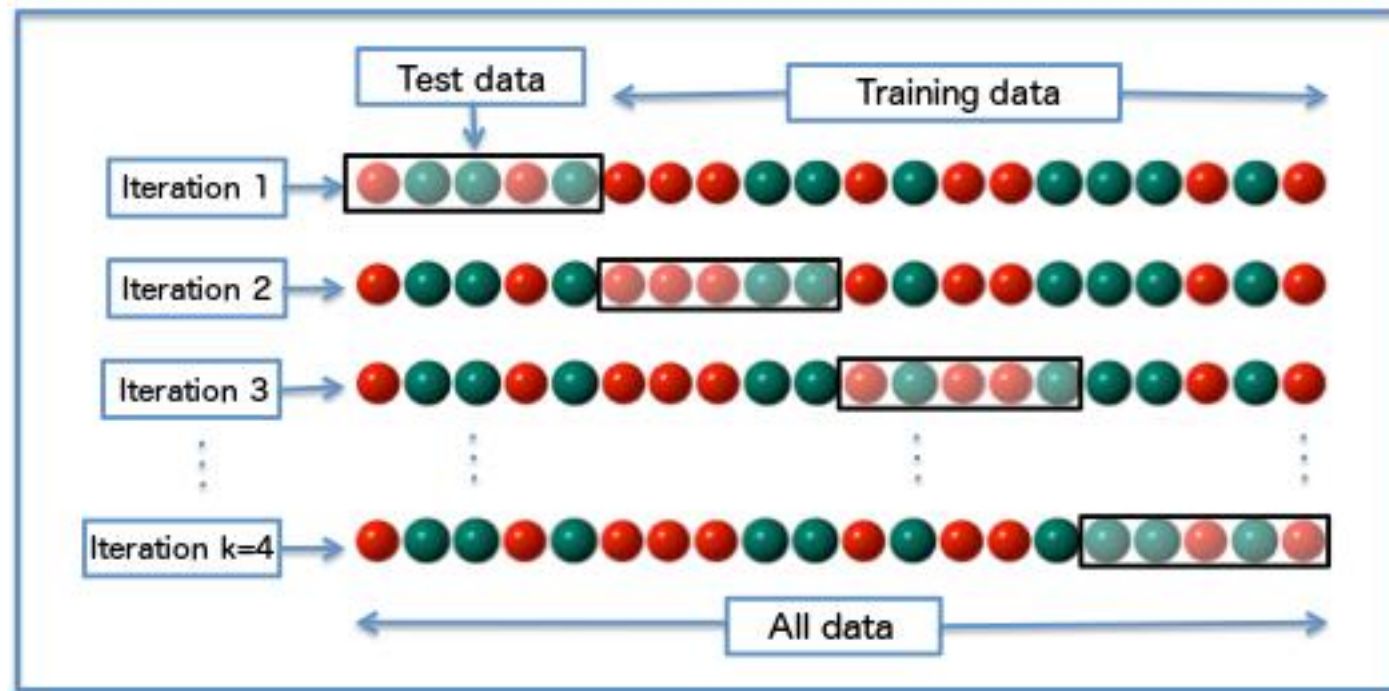
Attention, vous devez créer vos ensembles d'entraînement et de test de manière aléatoire !



## 3. Evaluation

- ▶ Apprenez vos paramètres sur l'ensemble d'apprentissage contenant 70 % de l'ensemble de données (en minimisant l'erreur d'apprentissage  $J_{train}(W)$ ).
- ▶ Calculer l'erreur de l'ensemble de test contenant 30 % de l'ensemble de données. (évaluation sur  $J_{test}(W)$ )
- ▶ Grâce aux ensembles d'entraînement et de test, nous pouvons observer comment le modèle se comporte sur de nouvelles données
- ▶ Mais nous avons un biais, si par hasard nous avons un ensemble de tests qui est facile à prédire, nous allons sous-estimer l'erreur, si au contraire nous avons un ensemble de données qui est difficile à prédire, nous allons surestimer l'erreur.

## 4. Cross-validation



C'est une meilleure façon de mesurer la performance d'un modèle.