Métriques de performances

Cours de machine learning avancé

Sommaire

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

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

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

- Need hard prediction
- How frequent our class prediction is correct.
- Accuracy between 0 and 1, the higher the better.

Accuracy = 0.99 !!

True labels	Predictions
10 cats	0 cats
990 dogs	1000 dogs

3. Cross-entropy loss or log loss

Binary:

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

Multiclass:

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

4. Confusion matrix

		Predicted label	
		$\hat{y} = 0$	$\hat{y} = 1$
True label	$y = 0$	6 TN	9 FP
	$y = 1$	1 FN	10 TP

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

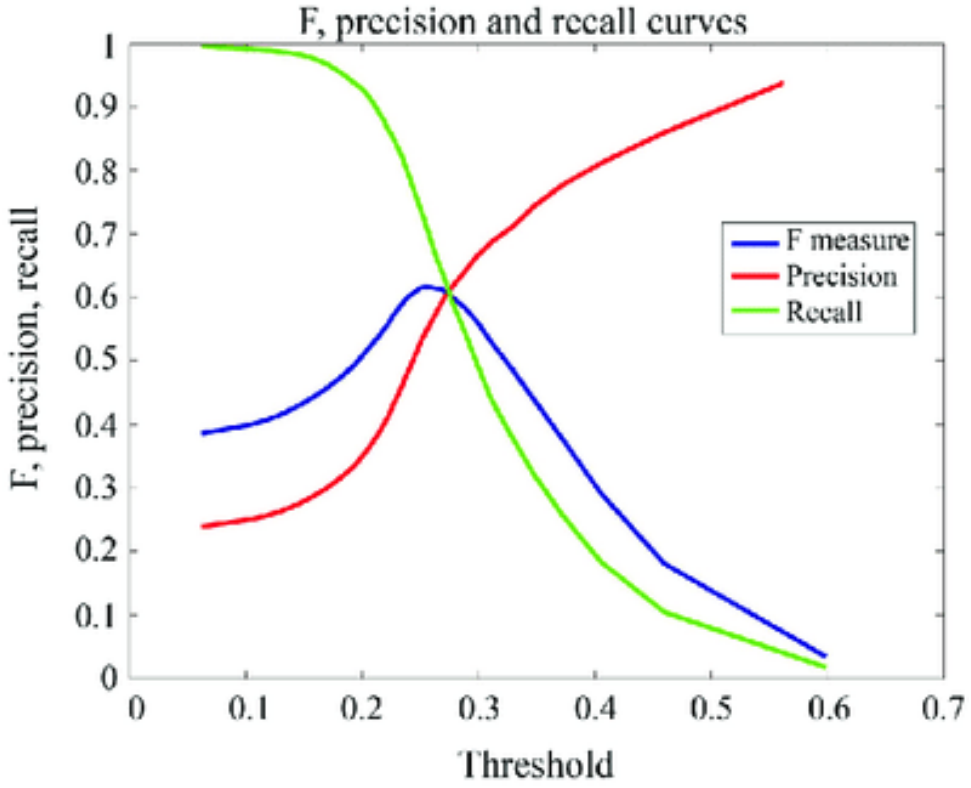
FPR = $\frac{FP}{FP+TN}$ = 1 - Specificity

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

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



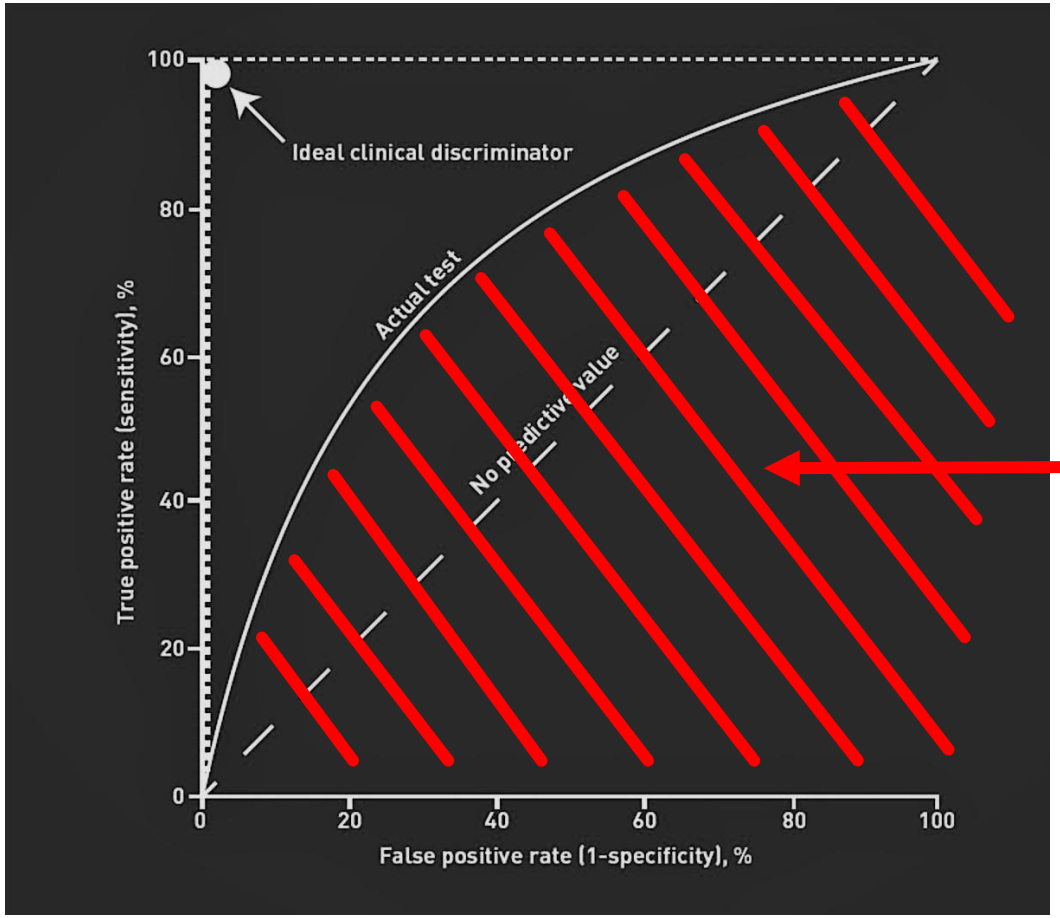
Recall, Sensitivity = $\frac{TP}{TP+FN}$

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

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

$$F_1 = F = 2\,\frac{precision\,.\,recall}{precision + recall}$$

6. AUC & ROC curve



Area Under Curve (AUC)

Sommaire

1. Generalization

- In machine learning the aim is to create an algorithm which has a great performances with new data. We call that concept generalization power. To measure the generalization of our model, we will predict data that our algorithm did not see during its training and see how it performs on this set.

2. Train and test

	Size (x_1)	Nb of room (x_2)	Year (x_3)	Price (y)
1	70	3	2010	460
2	40	3	2015	232
3	45	4	1990	315
4	12	2	2017	178
…	…	…	…	…
m-2	60	3	2010	390
m-1	35	2	1994	300
m	25	1	2005	240

Training set (70%)

Test set (30%)

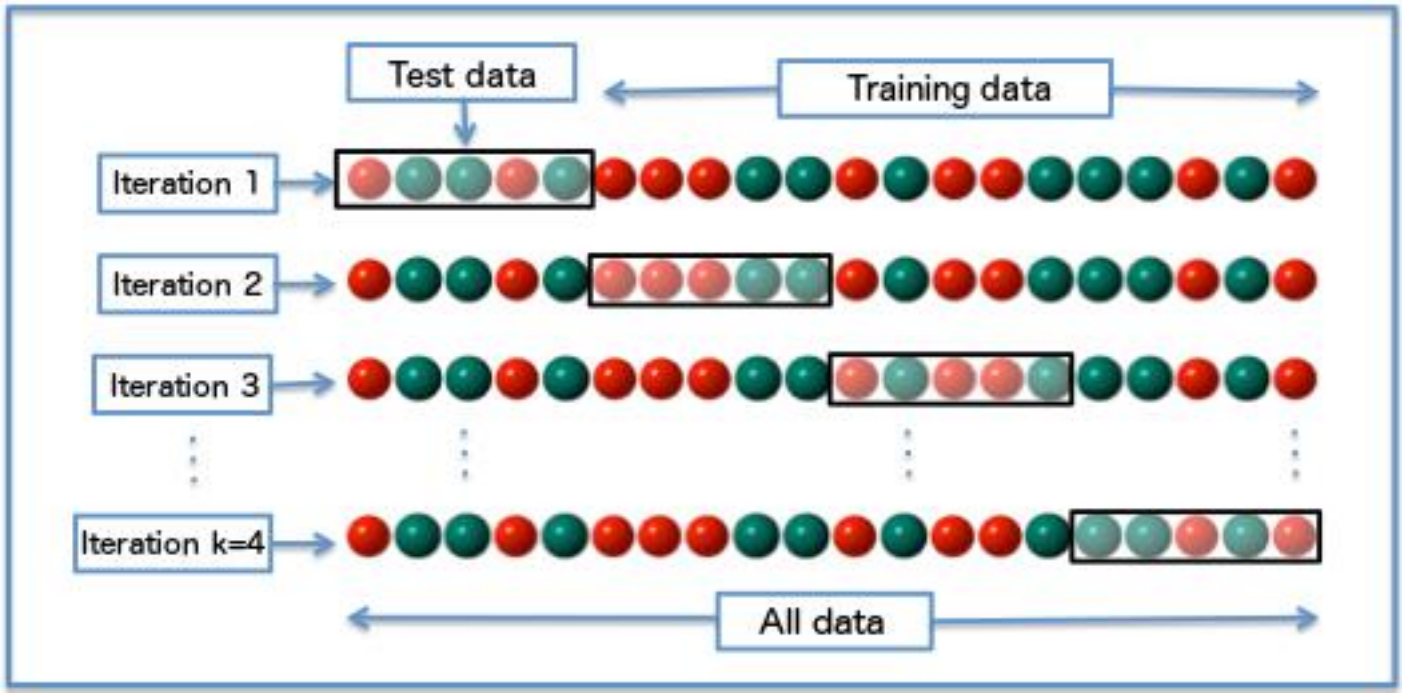
Training data prediction of the price of a house

Be careful, you have to create your training and your test set **randomly** !!

3. Evaluation

- Learn you parameters on the training set containing 70% of the dataset (minimizing training error $J_{train}(W)$).
- Compute test set error containing 30% of the dataset. (evaluate $J_{test}(W)$)
- Thanks to the training and testing set, we see how the model perform with new data. We can determine his level of generalization.
- But we have a bias, if by chance we have a test set that is easy to predict we will underestimate the error, if on the contrary we have a data set that is hard to predict we will overestimate the error.

4. Cross-validation



This is a better way to measure the performance of a model.