

Techniques d'Apprentissage Artificiel

Mesures de performance

Matrice de confusion (Confusion Matrix)

Une **Confusion Matrix** = matrice de confusion = tableau de contingence.

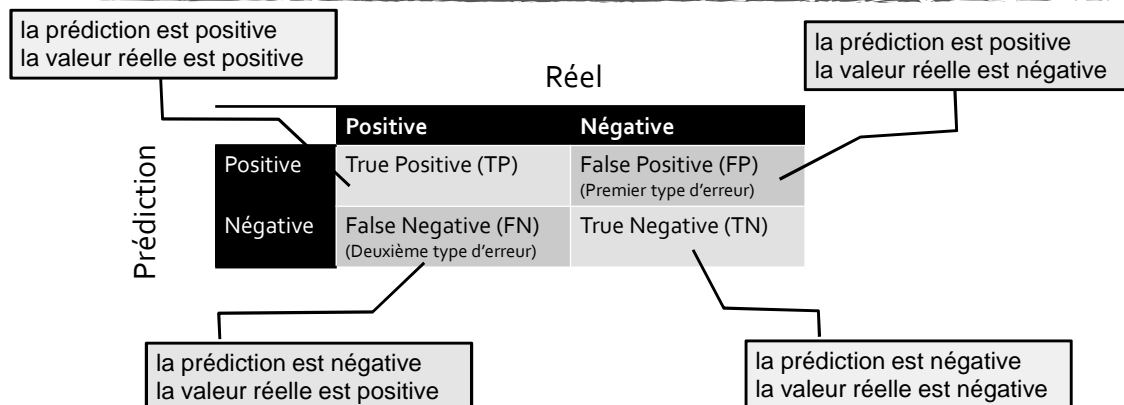
- permet de mesurer les performances d'un modèle de classification
- résume les résultats de prédictions sur un problème de classification. Les prédictions correctes et incorrectes selon les classes sont numériquement déterminés.
- permet la comparaison des résultats du modèle avec les valeurs réelles
- permet de comprendre la confusion du modèle de classification lorsqu'il effectue des prédictions.
- aide à reconnaître les erreurs commises par le modèle et leurs types

Matrice de confusion

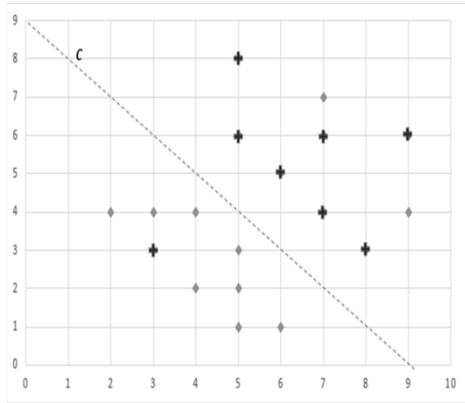
- la Confusion Matrix s'affiche sous forme d'un tableau 2x2.
- Le nombre de lignes et de colonnes varie en fonction du nombre de classes.
- Les **colonnes** correspondent aux **valeurs réelles**
- Les **lignes** indiquent les **valeurs prédites**.
- Pour interpréter le tableau, il suffit d'examiner l'intersection des lignes et des colonnes.

		True Class			
		A	B	C	D
Predicted Class	A				
	B				
	C				
	D				

Matrice de confusion



Exemple



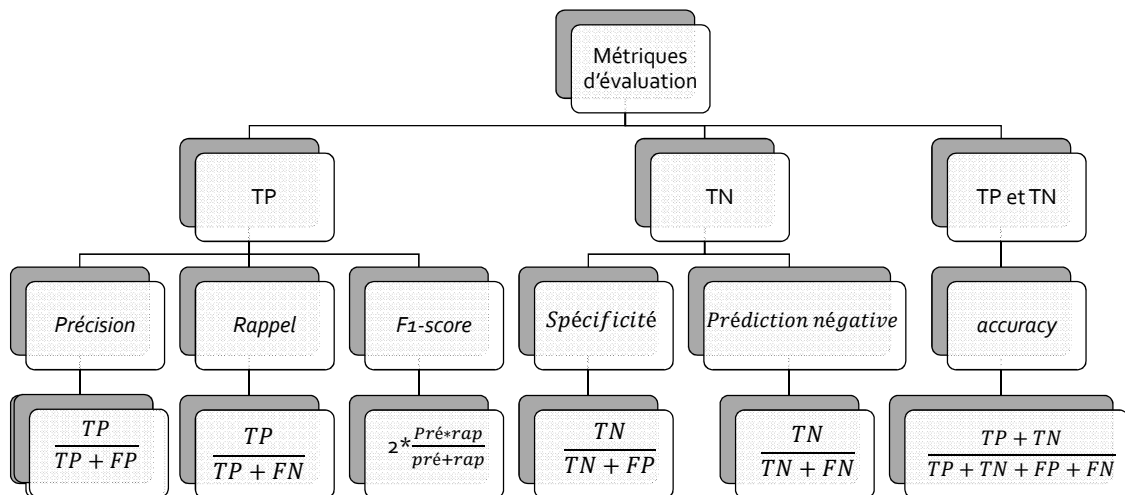
Le modèle classe correctement le point (5,8)

	P	N
P	TP=7	FP=2
N	FN=1	TN=8

Matrice de confusion et Métriques d'évaluation

		Réel		
		Positive	Négative	
Prédiction	Positive	(TP)	(FP)	$Précision = \frac{TP}{TP + FP}$
	Négative	(FN)	(TN)	$Prédiction\ négative = \frac{TN}{TN + FN}$
		$Rappel = \frac{TP}{TP + FN}$	$Spécificité = \frac{TN}{TN + FP}$	$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
	TP+FN=nombre de positifs réels	TN+FP=nombre de négatifs réels		

Métriques d'évaluation



Métriques d'évaluation

La précision

- permet de répondre à la question: "Quelle proportion d'identifications positives était effectivement correcte?"
- Quant il prédit un cas est positif sa prédiction est précise selon la valeur de la précision

$$\text{Précision} = \frac{\text{nbr d'individus correctement attribués à la classe } i}{\text{nbre d'individus attribués à la classe } i} = \frac{TP}{TP+FP}$$

Métriques d'évaluation

Rappel

- Le rappel ou sensibilité ou taux de vrais positifs (TVP)/ en anglais recall, Sensitivity, True Positive Rate (TPR)
- correspond au pourcentage d'exemples positifs qu'un modèle a automatiquement classé parmi tous les exemples positifs.
- Il permet de répondre à la question suivante: "Quelle proportion de résultats positifs réels a été identifiée correctement?"

$$\text{Rappel} = \frac{\text{nbr d'individus correctement attribués à la classe } i}{\text{nbre d'individus appartenant à la classe } i} = \frac{TP}{TP + FN}$$

Métriques d'évaluation (TP)

F1-score

- est une moyenne harmonique de la précision et du rappel

$$F1 - score = 2 * \frac{\text{Précision} * \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

Métriques d'évaluation

Spécificité ()

- Spécificité ou sélectivité ou taux de vrais négatifs (TN) / en anglais specificity, Selectivity, True Negative Rate (TNR)
- s'intéresse au taux d'exemples négatifs classés par le modèle
- mesure le ratio entre le nombre de prédictions négatives correctes et le nombre total de cas réels négatifs.
- La spécificité apparaît complémentaire au rappel

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

FPR (False positif rate)
Taux de faux positifs

Métriques d'évaluation

Prédiction négative ()

- en anglais False Negative Rate (FNR)
- s'intéresse au taux d'exemples négatifs classés par le modèle)

$$\text{Spécificité} = \frac{TN}{TN + FP}$$

Métriques d'évaluation

Taux de reconnaissance / Accuracy

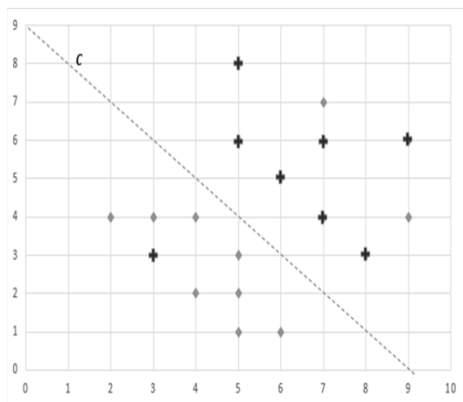
• Taux de reconnaissance ou Justesse ou Taux de réussite ou encore taux de prédiction; accuracy en anglais.

accuracy = nombre de prédictions correctes / nombre total de prédictions données *)

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Le taux d'erreur = 1-accuracy

Exemple



+ : personne malade

	P	N
P	TP=7	FP=2
N	FN=1	TN=8

Exemple

Quand le modèle prédit qu'une personne est malade, sa précision est juste dans 77% des cas

	malade	Non malade	
malade	(TP=7)	(FP=2)	$Précision = \frac{TP}{TP+FP} = \frac{7}{9} = 0,77$
Non malade	(FN=1)	(TN=8)	$Prédiction négative = \frac{TN}{TN+FN} = \frac{8}{9} = 0,88$
	$Rappel = \frac{TP}{TP+FN} = \frac{7}{8} = 0,875$	$Spécificité = \frac{TN}{TN+FP} = \frac{8}{10} = 0,8$	$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{15}{18} = 0,833$

Exemple

	malade	Non malade	
malade	(TP=7)	(FP=2)	$Précision = \frac{TP}{TP+FP} = \frac{7}{9} = 0,77$
Non malade	(FN=1)	(TN=8)	$Prédiction négative = \frac{TN}{TN+FN} = \frac{8}{9} = 0,88$
	$Rappel = \frac{TP}{TP+FN} = \frac{7}{8} = 0,875$	$Spécificité = \frac{TN}{TN+FP} = \frac{8}{10} = 0,8$	$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{15}{18} = 0,833$

Le modèle identifie correctement plus de 87% des personnes malades

Exemple

	malade	Non malade	
malade	(TP=7)	(FP=2)	$Précision = \frac{TP}{TP+FP} = \frac{7}{9} = 0,77$
Non malade	(FN=1)	(TN=8)	$Prédiction négative = \frac{TN}{TN+FN} = \frac{8}{9} = 0,88$
	$Rappel = \frac{TP}{TP+FN} = \frac{7}{8} = 0,875$	$Spécificité = \frac{TN}{TN+FP} = \frac{8}{10} = 0,8$	$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{15}{18} = 0,833$

80% sont des vrais négatifs

Exercice

Considérant le problème de classification défini par le Tableau 1. Pour le résoudre on utilise un perceptron à un neurone ayant :

- La fonction d'activation seuil
- $\eta=1$ et
- $(w_0^0, w_1^0, S^0) = (-0.182, 0.379, -0.493)$

1. Entraîner le perceptron de sorte qu'il résout le problème de classification posé. Quelles sont les valeurs de w_0 , w_1 et S obtenues.

2. Vérifier la performance de votre perceptron en calculant sa précision (accuracy) pour les données de tests du Tableau 2.

N°	x_0	x_1	t
1	-1	1	1
2	-1	-1	1
3	0	0	0
4	1	0	0

N°	x_0	x_1	t
1	-1	0	0
2	1	1	1
3	0	1	0
4	1	-1	0

Solution

$$w_0 = \mathbf{-1.182}$$

$$w_1 = \mathbf{-0.621}$$

$$S = \mathbf{-0.493}$$

$$\text{accuracy} = 50\% = 0.5$$