

But : quantifier la qualité d'un classificateur ou d'une méthode

Classificateur

= modèle

= méthode + paramètres pour l'apprentissage + jeu d'apprentissage

Taux d'erreurs observé sur un jeu de données de test (ou de validation)

Différentes approches selon

- la quantité de données disponibles (dont la classe est connue)
- l'objectif : comparer les méthodes et/ou leur paramètre, ou construire un modèle pour son utilisation

## Classification en 2 classes : Positive (P) et Negative (N)

P	Positive	N	Negative
PP	Predicted Positive	PN	Predicted Negative
TP	True Positive	TN	True Negative
FP	False Positive	FN	False Negative

Le plus simple → taux d'erreurs =  $(FP + FN) / (P + N)$   
 correspond à  $1 - accuracy = (TP + TN) / (P + N)$

**Sensibilité** (*sensitivity, Recall, TPR, ...*) =  $TP / P$

**Spécificité** =  $TN / N$

*Precision* =  $TP / (TP + FP) = TP / PP = PPV$  (*Positive Predictive Value*) ;  $NPV = TN / PN$

$$F_1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$

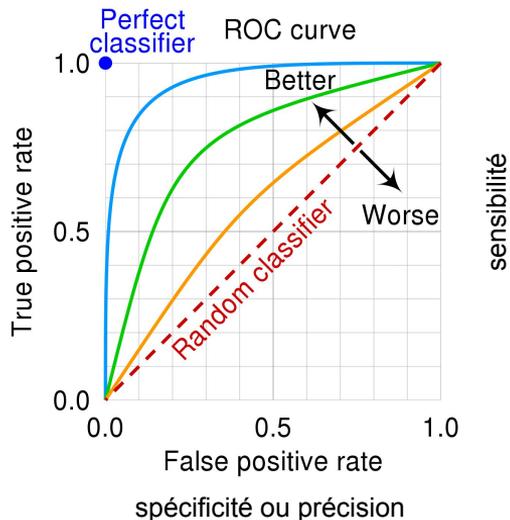
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

		réalité		
		P	N	
prédiction	PP	TP	FP	valeur prédictive positive (PPV)
	PN	FN	TN	valeur prédictive négative (NPV)
		sensibilité	spécificité	

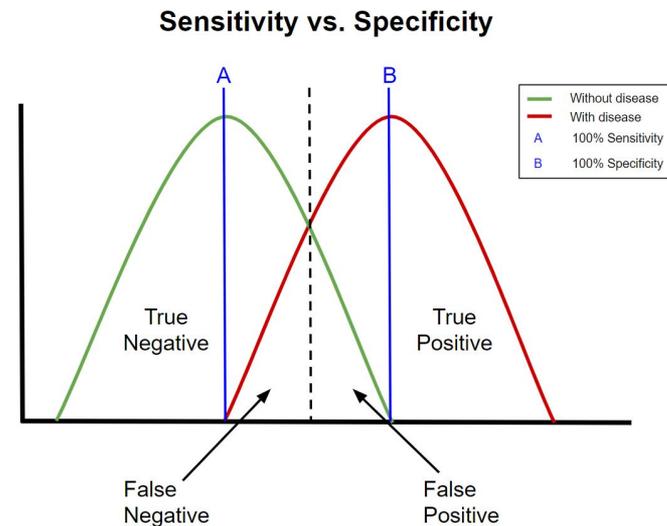
autres : [https://en.wikipedia.org/wiki/Sensitivity\\_and\\_specificity](https://en.wikipedia.org/wiki/Sensitivity_and_specificity)

## Courbe ROC + AUC

- 2 classes (P et V)
  - Principe : variation d'un seuil
  - ex : score pour la présence d'un domaine
  - $\alpha$  pour un test statistique
- ROC pour *Receiver operating characteristic* (= courbe sensibilité vs. spécificité ou précision)
- AUC pour *Area Under the Curve* (= aire sous la courbe)



By cmglee, MartinThoma - Roc-draft-xxcd-style.svg, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=109730045>



By Blue64701 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=97286749>

## Classes multiples (plus que 2)

3 classes (ou plus)

$$\text{Precision} = TP / (TP+FP)$$

$$\text{Recall} = TP/P$$

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

	Precision	Recall	F1-score	n
banane	0.64	0.5	0.56	14
citron	0.45	0.56	0.5	17
orange	0.57	0.53	0.55	15

		réalité		
		banane	citron	orange
prédiction	banane	7	3	1
	citron	5	9	6
	orange	2	4	8

bananes :

$$TP = 7 ; TN = 9 + 6 + 4 + 8 = 27$$

$$FP = 3 + 1 = 4 ; FN = 5 + 2 = 7$$

$$\text{Precision} = 7 / (7+4) = 0.64$$

$$\text{Recall} = 7 / (7+7) = 0.5$$

## Partitionnement

- utilisation de jeux indépendants :
  - apprentissage (2/3)
  - test (1/3)

## Validation croisée

- diviser les données en  $k$  partitions
- utiliser  $k-1$  partitions pour l'apprentissage et la dernière pour le test
- précision = nombre d'objets bien classés lors des  $k$  itérations / nombre d'objets
- leave-one-out (validation croisée avec  $k = s$ )

## Bootstrapping

- tirage aléatoire avec remise des objets constituant le jeu d'apprentissage