

Science des données III : module 2



Validation croisée et interprétation

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Laboratoire d'Écologie numérique

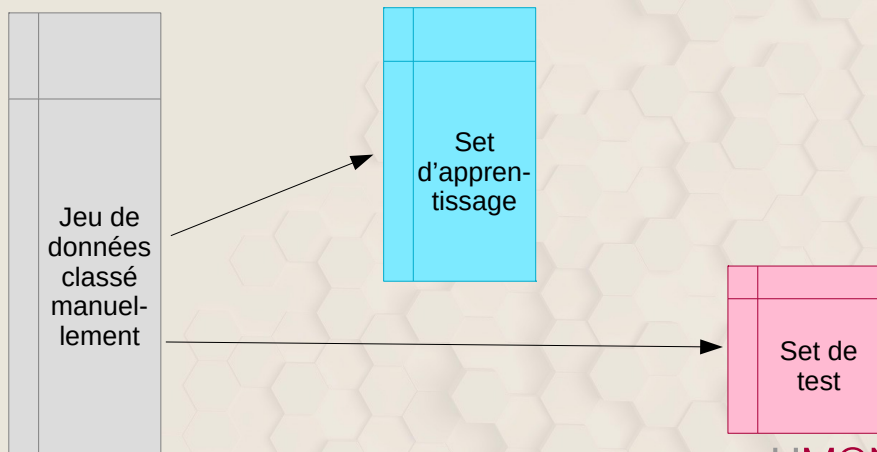


<https://wp.sciviews.org>
sdd@sciviews.org

Validation croisée

La validation croisée

L'obligation de séparer ses précieuses données en set d'apprentissage et set de test est très contraignante.



La validation croisée (2)

- La **validation croisée** est une technique générale qui permet d'utiliser toutes les données en apprentissage et en test, mais **pas simultanément**
- L'évaluation des performances en non biaisée

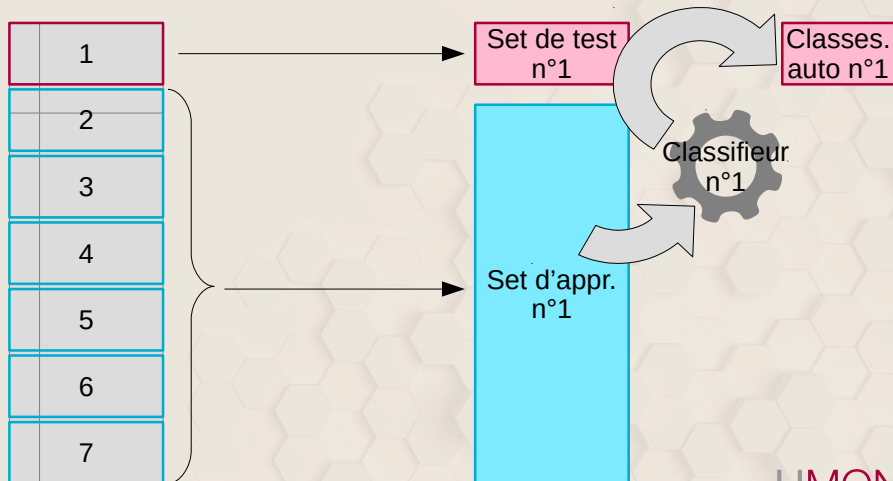
La validation croisée - principe

Séparation aléatoire en k sous-ensembles de mêmes effectifs
(exemple : jeu de données classé manuellement séparé en $k = 7$ blocs).

	1
	2
	3
	4
	5
	6
	7

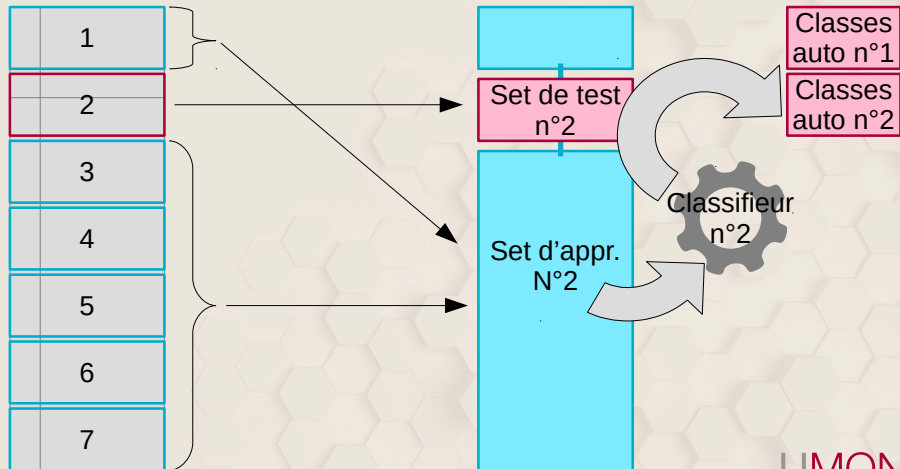
La validation croisée - principe

Etape 1 : le premier bloc est set de test, le reste set d'apprentissage. Classification du set de test.



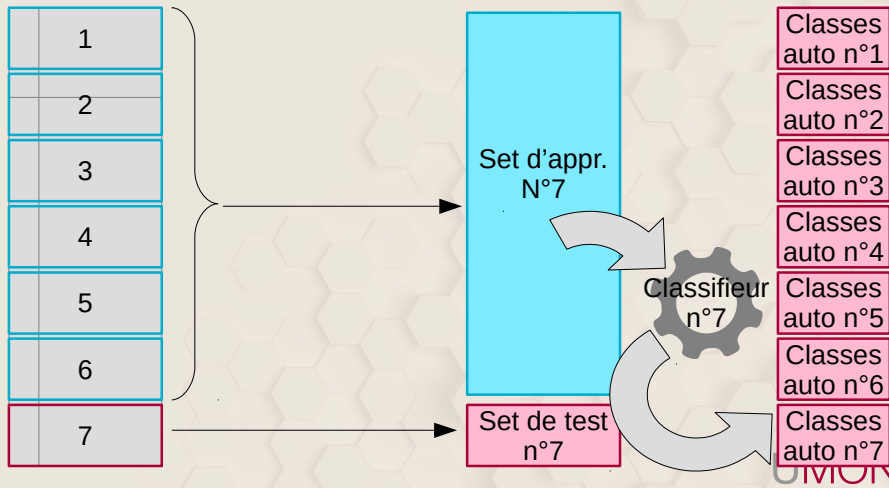
La validation croisée - principe

Etape 2 : le second bloc est set de test, le reste set d'apprentissage. Pooler la classification du set de test avec celle issue de l'étape précédente.



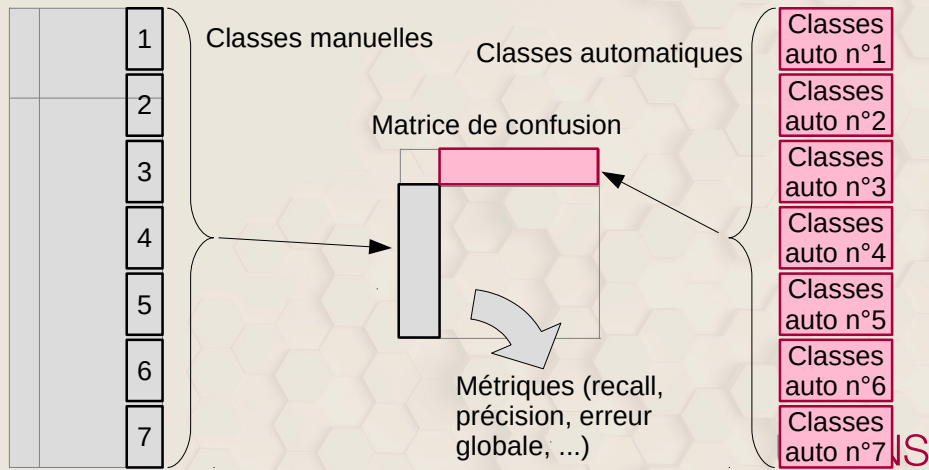
La validation croisée - principe

Etape 7 : à ce stage, k classifieurs ont été utilisés pour classer au final tous les individus du jeu de données initial.



La validation croisée - principe

Etape finale : les classes manuelles et automatiques de l'ensemble sont croisées dans la matrice de confusion, et les métriques sont calculées pour quantifier les performances.



Interprétation

Interprétation

...

Utilisation des métriques

...