

cible $\in \{+ : 1, - : 0\}$

Base initiale (DATASET)

Variables prédictives (X1, ..., Xp)	Cible (Y)	statut
		app
		app
		app
		app
		app
		app
		app
		app
		app
		app
		test
		test
		test
		test
		test

statut = app

Base d'apprentissage (TRAIN SET)

Variables prédictives (X1, ..., Xp)	Cible (Y)
(1)	

statut = test

Variables prédictives (X1, ..., Xp)	Cible (Y)
(3)	

Base de test (TEST SET)

Subdivision des données (0) en échantillons d'apprentissage (1) et de test (3).

Modélisation (2)

(2)

Modélisation –
Régression logistique :

$$LOGIT = a_0 + a_1x_1 + a_2x_2 + \dots$$

$$score_+ = \frac{1}{1 + e^{-LOGIT}}$$

Prédiction

Application du modèle **sur l'échantillon test**. Calcul des scores d'affectation

(4)

DASK : Déclenchement des calculs ici !

Cible (Y)	Score+

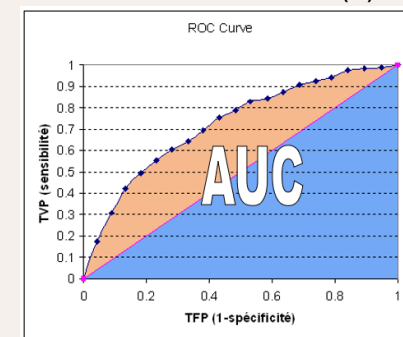
Mesure des performances

Confrontation entre valeurs observées de Y et les scores d'appartenance à la modalité cible (+)

(5)

Courbe ROC

(Voir [Support de cours](#))



Evaluation du modèle
prédictif (4 et 5)