

Apprentissage Statistique

Année 2018-2019

Exercices supplémentaires

A rendre avant le 28/12/2018

Exercice 1. On dispose des données suivantes :

X_1	X_2	Y
0.00	1.00	0.00
3.00	3.00	0.00
1.00	2.00	0.00
3.00	5.00	1.00
-6.00	-8.00	1.00
-1.00	-2.00	1.00
-1.00	-2.00	1.00
1.00	2.00	1.00

Et l'on souhaite prédire la classe Y pour un nouvel individu pour lequel on observe $X_1 = 0$ et $X_2 = -1$.

1. On souhaite appliquer la méthode des k plus proche voisins avec $k = 1$ directement sur les observations. En expliquant soigneusement toutes les étapes, prédire la classe du nouvel individu.
2. On se rend compte que les co-variables X_1 et X_2 n'ont pas tout à fait la même dispersion.
 - 2.1. Calculer la variance de chacune des variables X_1 et X_2 .
 - 2.2. Expliquer soigneusement pourquoi cela peut être problématique pour la méthode de classification des k plus proches voisins.
 - 2.3. Comment remédier à ce problème ? (indication : changer la distance)
 - 2.4. Prédire la classe du nouvel individu avec cette nouvelle méthode en prenant $k = 1$.

Exercice 2. Pour les données `australian_credit` vous proposerez (en motivant votre choix) puis mettrez en œuvre une méthode de classification supervisée. Vous prendrez soin d'expliquer les différentes étapes de calibration et d'évaluation de la qualité de la méthode choisie.