

Detection des points aberrants

Interpolaspline

Avril 2020

1 Méthode des distance

Dans cette partie, il est question de la détection des points aberrants en utilisant les calculs de distance.

1.1 Méthode des k plus proches voisins

Calculer pour chaque observation la distance au K plus proche voisin k-distance ; Ordonner les observations selon ces distances k-distance ; Les données aberrantes ont les plus grandes distances k-distance ; Les observations qui ont les n pourcent plus grandes distances k-distance sont des données aberrante, n étant un paramètre à fixer.

Dans un premier temps, nous avons crée un algorithme qui prend en entrée une liste d'abscisses et une autre d'ordonnées, un point abscisse et ordonnées et un entier k, retourne la k-distance du point donné en entrée la moyenne ses k petites distances.

Enfin, notre fonction KNN qui comme la précédente, prend 2 listes une d'abscisses et l'autre d'ordonnées, un entier k et un entier m comme indiqué au dessus, retourne 4 listes les deux premières(abscisses et ordonnées) contenant les valeurs de la liste qui ont les n pourcent plus grande k-distance qui représentent ici les valeurs aberrantes et les 2 dernières les abscisses et ordonnées des valeurs ici considérées comme non aberrantes.

1.2 Exemple d'exécutions des proches voisins

Pour cette première figure ci dessous nous effectuons le test pour $k = 3$ et un pourcentage de valeurs à rejeter de 45. Et une seconde pour $k = 5$ et $m = 25$. Sur les nombreux tests réalisés on remarque que plus k est petit et m grand la méthode donne plus de précisions sur les valeurs aberrantes

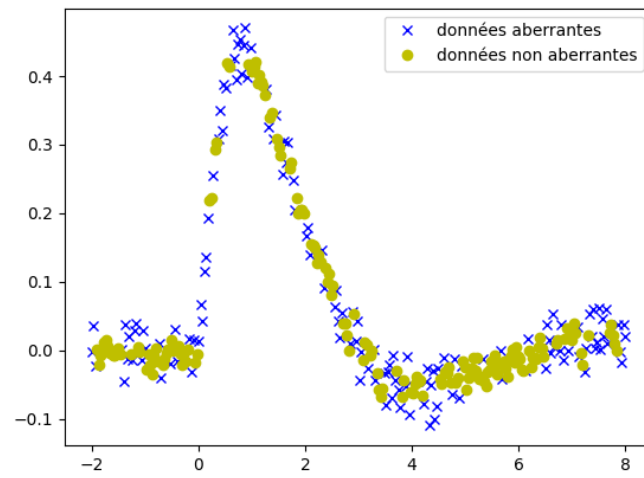


FIGURE 1 – Exemple du résultat $k = 3$ et $m = 45\%$

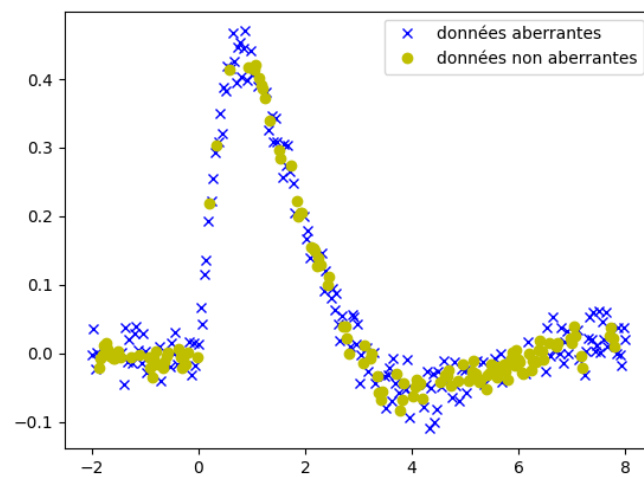


FIGURE 2 – Exemple du résultat $k = 5$ et $m = 25\%$