



Module 4 - Estimation non paramétrique

Exercices - Consignes

Exercice 1

1. Quelle est la différence entre les méthodes d'estimation de densité paramétrique et non-paramétrique ?
2. Décrivez brièvement la méthode d'estimation des k plus proches voisins.
3. Comment peut-on choisir la valeur du paramètre k dans une méthode de classification par les k plus proches voisins ? Soyez précis.

Exercice 2*

Soit un ensemble de données $D = \{4, 5, 5, 6, 12, 14, 15, 15, 16, 17\}$.

1. Utilisez une fenêtre de Parzen de largeur $h = 4$ et de dimension unitaire pour calculez la valeur de la densité $p(x)$ pour $x = 3$, $x = 10$, et $x = 15$.
2. Utilisez une fenêtre de Parzen de largeur $h = 5$ et de dimension unitaire pour calculez la valeur de la densité $p(x)$ pour $x = 3$, $x = 10$, et $x = 15$.

3. Comparez les résultats précédents (pour $h = 4$ et $h = 5$).

Exercice 3*

Soit un ensemble de données $D = \{4, 5, 5, 6, 12, 14, 15, 15, 16, 17\}$. En utilisant le logiciel R :

1. Représentez graphiquement les données de l'ensemble D .
2. Utilisez une fenêtre de Parzen de largeur $h = 4$ et de dimension unitaire pour calculez la valeur de la densité $p(x)$ pour $x = 3$, $x = 10$, et $x = 15$.
3. Utilisez une fenêtre de Parzen de largeur $h = 5$ et de dimension unitaire pour calculez la valeur de la densité $p(x)$ pour $x = 3$, $x = 10$, et $x = 15$.

Exercice 4*

Soit la base de données IRIS de R.

1. Téléchargez cette base de données.
2. Énumérez les variables et les classes de formes et identifiez le nombre d'échantillons de la base de données.
3. Visualisez les données des classes d'Iris en considérant les variables `Sepal.Length` et `Petal.Length`.
4. Formez un ensemble d'entraînement et un ensemble de test en utilisant un partition 80% pour l'entraînement et 20% pour le test.
5. Réalisez une classification à l'aide de l'estimateur des k -plus voisins en prenant $k = 3$ (`set.seed=147`). Donnez la matrice de confusion.
6. Réalisez une classification à l'aide de l'estimateur des k -plus voisins en prenant $k = 7$ (`set.seed=147`). Donnez la matrice de confusion.

7. Comparez les deux résultats précédents ($k = 3$ et $k = 7$).