



## Module 6 - Algorithmes de base en apprentissage machine

### Exercices - Consignes

Cette série de questions a pour objectif de manipuler les modèles de regroupement de données et de régression sur la base de données IRIS. Certaines questions nécessitent une exploration au delà des informations fournis dans le texte de référence du module 6. Dans ces cas, une recherche scientifique, via internet par exemple, est suggérée. Vous devez fournir, pour chaque question, les lignes de codes qui permettent de répondre à la question, s'il y a lieu.

NB : Les questions marquées par \* ne sont corrigées.

#### Exercices 1 :

Dans cet exercice, nous allons étudier les modèles de régression linéaires.

1. Téléchargez le contenu de cette base de données et décrivez la.
2. Représentez graphiquement la variation des variables `Sepal.Width` en fonction de `Sepal.Length`. Attribuez une couleur à chaque type d'espèce.
3. Identifiez la relation entre `Sepal.Width` et `Sepal.Length` en utilisant la corrélation de Pearson.
4. En considérant la variable `Sepal.Width` dépendante et la variable `Sepal.Length` comme variable indépendante, identifiez une relation de régression entre ces deux variables.

#### Exercices 2\* :

Dans cet exercice, nous allons étudier les modèles de regroupement.

1. Téléchargez le contenu de la base de données et décrivez la.
2. Représentez graphiquement la variation des variables des sépales (`Sepal.Width` en fonction de `Sepal.Length`). Attribuez une couleur à chaque type d'espèce.
3. En utilisant l'algorithme  $k$ -moyenne, réalisez un regroupement en 3 groupes en se basant sur les mesures des sépales.
  - (a) Déterminez le nombre d'échantillons par groupe.
  - (b) Représentez la silhouette
4. En utilisant l'algorithme  $k$ -moyenne, réalisez un regroupement en 4 groupes en se basant sur les mesures des sépales.
  - (a) Déterminez le nombre d'échantillons par groupe.
  - (b) Représentez la silhouette
  - (c) Comparez les résultats pour 3 et 4 groupes
5. En utilisant l'algorithme  $k$ -moyenne, réalisez un regroupement en 3 groupes en se basant sur les mesures des sépales et des pétales.
  - (a) Déterminez le nombre d'échantillons par groupe.
  - (b) Représentez la silhouette
  - (c) Comparez les résultats avec ceux de la question 3.