

## Exercices

### Cours 1 : Les fondements de l'apprentissage machine

Date : 3 janvier 2022

Soit l'ensemble des données `Iris` qui comprend 150 observations de fleurs d'iris décrites par la longueur et la largeur des sépales et des pétales. Trois espèces différentes sont incluses `Iris setosa`, `Iris versicolor` et `Iris virginica`

#### Exercice 1 : Machine à support vectoriels

1. Téléchargez le contenu de la base de données (`iris = datasets.load_iris()`)
2. Répartir l'ensemble de données en entraînement (50% de l'ensemble des données) et test (50% de l'ensemble des données)
3. Représentez la dispersion de la variable `Sepal width` en fonction de `Sepal length` pour l'ensemble des données d'entraînement et l'ensemble des données de test. Utilisez des `marker` différents.
4. On se limite aux variables `Sepal width` et `Sepal length`. Entraînez un SVM linéaire sur l'ensemble des données d'entraînement. Utilisez la librairie `svm.LinearSVC`. Fixez la valeur du paramètre de régularisation à 1, `random_state = 10`.
5. Déterminez le taux de classification.
6. Déterminez la matrice de confusion et la matrice de confusion normalisée.
7. Représentez graphiquement les frontières de décision.
8. On se limite toujours aux variables `Sepal width` et `Sepal length`. Entraînez un SVM linéaire sur l'ensemble des données d'entraînement. Utilisez la librairie `svm.SVC`. Fixez la valeur du paramètre de régularisation à 1, `random_state = 10`.
9. Comparez les deux modèles (obtenus en utilisant `svm.LinearSVC` et `svm.SVC`)

10. En utilisant les 4 caractéristiques, entraînez un SVM linéaire sur l'ensemble des données d'entraînement. (Fixez la valeur du paramètre de régularisation à 1, `random_state = 10`).
11. Déterminez le taux de classification.
12. Comparez les résultats des questions 7 et 10.

## **Exercice 2 : Arbres de décision**

1. Répartir l'ensemble de données en entraînement (70% de l'ensemble des données) et test (30% de l'ensemble des données)
2. Entraînez un arbre de décision sur l'ensemble des données d'entraînement de profondeur 3.
3. Déterminez le taux de classification et la matrice de confusion
4. Entraînez un arbre de décision sur l'ensemble des données d'entraînement avec un nombre maximum de feuille égal à 20.
5. Déterminez le taux de classification et la matrice de confusion
6. Répartir l'ensemble de données en entraînement (5% de l'ensemble des données) et test (95% de l'ensemble des données). Calculez le taux de classification pour les profondeurs 1 à 7.
7. Répartir l'ensemble de données en entraînement (5% de l'ensemble des données) et test (95% de l'ensemble des données). Calculez le taux de classification pour un nombre total de feuilles de 2, 3, 5, 10, 15, 20.
8. Déterminez les valeurs optimales de profondeur et de nombre de feuilles en utilisant la librairie `GridSearchCV` et en se basant sur le taux de classification.