

Devoir en Machine Learning

(Durée : 1h30)

Exercice 1: Régression Multiple

On étudie la relation entre le revenu mensuel (Y, en milliers d'euros) et deux variables explicatives pour quatre employés dans une entreprise

- X1 : le nombre d'années d'expérience,
- X2 : le nombre d'heures de formation suivies.

Les données pour quatre employés sont les suivantes:

Employés	Y	X1	X2
1	12	1	2
2	18	2	3
3	29	3	4
4	35	4	6

$$\text{Avec } (X^T X)^{-1} = \begin{pmatrix} 2,3333 & 1,6666 & -1,6666 \\ 1,6666 & 5,8333 & -4,3333 \\ -1,6666 & -4,3333 & 3,3333 \end{pmatrix}$$

- 1) Estimez les paramètres du modèle.
- 2) Calculez la somme des carrés totale (SCT), la somme des carrés expliquée (SCE) et la somme des carrés résiduelle (SCR).
- 3) Calculez le coefficient de détermination R^2 et interprétez le résultat.
- 4) Calculez les coefficients de corrélation entre Y, X1 et X2, et commentez les relations entre les variables.
- 5) Testez le modèle globalement et interprétez les résultats.

Exercice 2: Classification K-means

L'objectif est de réaliser une classification en utilisant la bibliothèque Scikit-Learn afin de diviser un jeu de données en trois groupes

- 1) Donnez le code permettant de charger les bibliothèques nécessaires pour effectuer la classification K-means
- 2) Donnez le code permettant d'appliquer la méthode K-means pour effectuer la classification
- 3) Donnez le code permettant d'afficher les groupes obtenus et les centres des clusters.
- 4) Donnez le code permettant de visualiser les résultats à l'aide d'un graphique.