## Activité pratique 1: Régression Linéaire et et Techniques de Régularisation

**Objectif :** Dans ce TP, vous allez modéliser la quantité d'expression de l'antigène associée à la détection du cancer de la prostate (lpsa) en fonction des différentes variables explicatives, en utilisant la régression linéaire et les techniques de régularisation.

1. Importer les bibliothèques Python suivantes : pandas, numpy, matplotlib, seaborn, scikit-learn, etc., et télécharger la base de données prostate_dataset.txt.
2. Supprimer les colonnes "col" et "train" du jeu de données.
3. Nettoyer les données si nécessaire (conversions de types, traitement des données manquantes, des données aberrantes,….) et réaliser une analyse exploratoire des données comprenant :
    - Des statistiques descriptives uni-variées et bi-variées.
    - La visualisation des données (ex : histogrammes, graphiques en nuage de points, etc.).
4. Standardiser les données explicatives.
5. Séparer les variables explicatives de la variable à prédire (psa).
6. Diviser le jeu de données en deux ensembles : un ensemble d'entraînement et un ensemble de test.
7. Entraîner un modèle de régression linéaire et évaluer sa performance sur l'ensemble de test.
8. Entraîner un modèle de régression Ridge avec un coefficient de régularisation $\lambda=2.14$ et évaluer sa performance sur l'ensemble de test.
9. Utiliser la validation croisée pour déterminer la valeur optimale de $\lambda$. Évaluer la performance du nouveau modèle.
10. Entraîner un modèle de régression Lasso avec $\lambda=0.08$ et évaluer sa performance sur l'ensemble de test.
11. Utiliser la validation croisée pour déterminer la valeur optimale de $\lambda$. Évaluer la performance du nouveau modèle.
12. Entraîner un modèle de régression Elastic Net avec les valeurs optimales de $\lambda$ et de $\alpha$, puis évaluer sa performance sur l'ensemble de test.
13. Intepréter les résultats trouvés

**Activité pratique 1: Régression Linéaire et et Techniques de Régularisation**

14. À l'aide du modèle choisi, prédire la quantité d'expression de l'antigène (lpsa) pour un patient avec les caractéristiques suivantes:

| lcavol | lweight | age | lbph | svi | lcp | gleason | Pgg45 |
|---|---|---|---|---|---|---|---|
| 2.8 | 3 | 70 | -1.4 | 1 | 1.5 | 7 | 60 |

15. Sauvegarder le modèle final dans un format appropri