

## TP N°2

### Apprentissage Supervisé et Régression

Master 1 (STIC & I2A)

Matière : Apprentissage Automatique

#### Exercice 1 : Régression linéaire simple

##### Objectif

Étudier et implémenter un modèle de **régression linéaire simple** à partir d'un jeu de données fourni, en utilisant :

- La méthode analytique des **moindres carrés**
- Une bibliothèque d'apprentissage automatique

##### Données

On considère le jeu de données suivant :

##### Travail demandé

1. Initialiser l'environnement de travail Python.
2. Représenter graphiquement les données sur un plan 2D.
3. Calculer les paramètres du modèle de régression linéaire en utilisant la méthode analytique des moindres carrés.
4. Afficher les paramètres obtenus (intercept et coefficient).
5. Prédire les valeurs de sortie pour :
  - $x = -1.5$
  - $x = 2$
6. Tracer la droite de régression superposée aux données d'apprentissage.
7. Comparer les résultats obtenus avec ceux fournis par une bibliothèque d'apprentissage automatique.

| x     | y     |
|-------|-------|
| 0.86  | 2.49  |
| 0.09  | 0.83  |
| -0.85 | -0.25 |
| 0.87  | 3.10  |
| -0.44 | 0.87  |
| -0.43 | 0.02  |
| -1.10 | -0.12 |
| 0.40  | 1.81  |
| -0.96 | -0.83 |
| 0.17  | 0.43  |

#### Exercice 2 : Régression polynomiale (degré 2)

Implémenter un modèle de **régression quadratique** (polynomiale de degré 2) sur le même jeu de données.

##### Travail demandé

1. Transformer les variables d'entrée pour obtenir les caractéristiques polynomiales de degré 2.
2. Calculer les paramètres du modèle de régression quadratique.
3. Afficher les coefficients du modèle obtenu.
4. Tracer :
  - Les données d'apprentissage
  - La courbe de régression quadratique
5. Comparer les performances entre :
  - La régression linéaire simple
  - La régression polynomiale
6. Discuter les notions suivantes :
  - Sous-apprentissage (underfitting)
  - Sur-apprentissage (overfitting)
  - Influence du degré du polynôme

$$\mathbf{X} = \begin{bmatrix} 0.75 & 0.86 & 1 \\ 0.01 & 0.09 & 1 \\ 0.73 & -0.85 & 1 \\ 0.76 & 0.87 & 1 \\ 0.19 & -0.44 & 1 \\ 0.18 & -0.43 & 1 \\ 1.22 & -1.10 & 1 \\ 0.16 & 0.40 & 1 \\ 0.93 & -0.96 & 1 \\ 0.03 & 0.17 & 1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 2.49 \\ 0.83 \\ -0.25 \\ 3.10 \\ 0.87 \\ 0.02 \\ -0.12 \\ 1.81 \\ -0.83 \\ 0.43 \end{bmatrix}$$

#### Exercice 3 : Classification Linéaire

## ◆ Partie : Classification Linéaire

Dans ce travail pratique, vous allez :

- Mettre en place un **problème de classification linéaire binaire**
- Utiliser le **jeu de données Iris**
- Étudier la **géométrie de la frontière de décision**
- Implémenter la classification linéaire :
  - À partir de la formule analytique vue en cours
  - En utilisant une bibliothèque d'apprentissage automatique

### 1. Présentation du jeu de données

Nous utilisons le célèbre jeu de données : **Iris dataset** Ce jeu de données contient :

- 150 échantillons
- 3 classes de fleurs :
  - Iris-Setosa
  - Iris-Versicolour
  - Iris-Virginica
- 4 caractéristiques numériques :
  - Longueur du sépale
  - Largeur du sépale
  - Longueur du pétale
  - Largeur du pétale

### 2. Simplification pour ce TP

Dans ce travail :

- On utilise uniquement les **deux premières classes** :
  - Setosa
  - Versicolour
- On utilise uniquement les **deux premières caractéristiques** :
  - Longueur du sépale
  - Largeur du sépale

On obtient ainsi un problème de **classification binaire en dimension 2**.

### 3. Modèle de classification linéaire

Le modèle étudié est :  $f(x) = w_0 + w_1x_1 + w_2x_2$

Règle de décision :

- Si  $f(x) \geq 0 \rightarrow$  classe 1
- Sinon  $\rightarrow$  classe 0

La frontière de décision est donnée par :  $w_0 + w_1x_1 + w_2x_2 = 0$

Il s'agit d'une **droite dans le plan**.

### 4. Travail demandé

#### Partie 1 : Exploration des données

1. Charger le jeu de données Iris.
2. Sélectionner uniquement :
  - Les 100 premiers échantillons (2 classes)
  - Les deux premières caractéristiques
3. Afficher les données dans un plan 2D.
4. Vérifier si les données sont linéairement séparables.

#### Partie 2 : Étude géométrique

1. Tracer plusieurs frontières de décision pour différents vecteurs de poids  $w_0, w_1, w_2$ .
2. Observer l'effet de la variation des paramètres sur :

- L'inclinaison de la droite
- La position de la frontière
- 3. Interpréter géométriquement le rôle :
  - Du biais  $w_0$
  - Des coefficients  $w_1$  et  $w_2$

### **Partie 3 : Implémentation depuis zéro**

Implémenter la classification linéaire à partir de la formule :  $W=(X^T X)^{-1} X^T Y$

Travail demandé :

1. Ajouter une colonne de 1 à la matrice des caractéristiques.
2. Calculer le vecteur de poids  $W$ .
3. Afficher les coefficients obtenus.
4. Tracer la frontière de décision.
5. Vérifier visuellement la qualité de la séparation.

### **Partie 4 : Utilisation d'une bibliothèque**

1. Résoudre le même problème à l'aide d'une bibliothèque d'apprentissage automatique.
2. Comparer les coefficients obtenus avec ceux calculés manuellement.
3. Tracer la nouvelle frontière de décision.
4. Vérifier que les résultats sont identiques.