

NOMS:  
CLASSE:  
EQUIPE:

### TP3 : Régression Polynomiale

Étudier la relation non linéaire entre certaines variables du jeu de données Boston Housing

Et utiliser la régression polynomiale pour :

- Mieux ajuster le modèle,
- Comprendre l'effet du degré du polynôme,
- Appliquer le BIC pour choisir le modèle optimal.

Dans ce TP, l'instruction « entraîner un modèle » doit être comprise comme le processus complet. Cela inclut implicitement la préparation des données.

#### **Partie 1 — Téléchargement et chargement du dataset**

1. Charger le dataset Boston Housing à l'aide de la bibliothèque `sklearn.datasets`.
2. Créer un `DataFrame` pandas à partir des données et afficher les 5 premières lignes.
3. Identifier la variable cible (MEDV) et la variable LSTAT

Quelle est la variable que nous cherchons à prédire ?

D'après toi, la relation LSTAT – MEDV semble-t-elle linéaire ?

#### **Partie 2 — Visualisation initiale**

1. Tracer un graphique en nuage de points entre LSTAT (axe X) et MEDV (axe Y).

Quelle forme générale observe-t-on ?

Une droite semble-t-elle adaptée ?

Si non, quelle autre forme de modèle serait plus appropriée ?

### Partie 3 — Régression linéaire simple

1. Entraîner un modèle de régression linéaire simple avec LSTAT comme variable explicative et MEDV comme variable cible.
2. Tracer la droite obtenue sur le nuage de points.
3. Calculer le MSE et le  $R^2$  du modèle.

Que représentent le **MSE** et le  **$R^2$**  ?

Quelle est la valeur du  $R^2$  obtenue ? Est-elle satisfaisante ?

Que nous indique visuellement la droite par rapport aux points ?

### Partie 4 — Régression polynomiale

1. Créer des variables polynomiales de degré 2 à partir de LSTAT.  
`poly = PolynomialFeatures(degree=2)`  
`X_poly = poly.fit_transform(X)`
2. Entraîner un modèle de régression linéaire sur ces nouvelles variables.
3. Tracer la courbe obtenue.
4. Calculer le MSE et le  $R^2$ .

Compare les valeurs du MSE et du  $R^2$  entre les modèles linéaire et polynomiale.

## Partie 5 — Étude du degré du polynôme

On veut maintenant faire varier le **degré du polynôme** (1, 2, 3, 4, 5, 6) et mesurer comment cela affecte le **MSE**, le **R<sup>2</sup>** et le **BIC**

Pour chaque degré :

- Créer les variables polynomiales ;
- Entraîner un modèle ;
- Calculer le **MSE**, le **R<sup>2</sup>** et le **BIC**.
- Rassembler les résultats dans un tableau.

Comment évoluent MSE et R<sup>2</sup> lorsque le degré augmente ?

Pourquoi le R<sup>2</sup> augmente toujours ?

Que signifie le paramètre  $k$  dans la formule du BIC ?

Quelle est la différence entre les deux termes du BIC :

Quel est le degré donnant le plus petit BIC ?

## Partie 6 — Visualisation du BIC

1. Tracer le BIC en fonction du degré du polynôme.
2. Identifier visuellement le degré où le BIC est minimal.

Que représente le minimum du BIC ?

Pourquoi le BIC diminue d'abord puis augmente ensuite ?

Quelle est la meilleure valeur du degré selon ce critère ?

## **Partie 7 — Visualisation des modèles**

1. Tracer sur le même graphique les courbes obtenues pour les degrés 1 à 6 (en couleurs différentes).
2. Comparer visuellement les formes.

Quels modèles semblent trop rigides ?

Quels modèles sont trop oscillants ?

Quel modèle te paraît le plus équilibré ?