# DM MLA

## Introduction

Les modèles de régression permettent d'expliquer une variable à partir d'une plusieurs variables explicatives.

### 1. Régression linéaire multiple

Dans un modèle de régression linéaire multiple, la variable cible est exprimée en fonction de plusieurs variables (explicatives)

$Y=a_0+a_1X_1+a_2X_2+ \ldots.. a_nX_n$

### 2. Régression logistique

La régression logistique est une méthode prédictive. Elle vise à construire un modèle permettant de prédire les valeurs prises par une variable cible qualitative binaire.

### 3. Régression logistique multinomiale

Si la variable cible possède plus de 2 modalités, on parle de régression logistique polyatomique ou multinomiale. Les variables explicatives sont qualitatives ou quantitatives (à coder).

### 4. Régression polynomiale

La régression polynomiale est une régression dans laquelle la relation entre la variable explicative et la variable expliquée est modélisée par un polynôme de degré n.

La régression linéaire est une régression polynomiale de degré 1.

Choix du degré du polynôme : faire attention au sur-apprentissage et au sous-apprentissage

*L'overfitting d'un modèle est une condition dans laquelle un modèle commence à décrire l'erreur aléatoire (le bruit) dans les données plutôt que les relations entre les variables. Ce problème se produit lorsque le modèle est trop complexe.*

*Dans l'autre sens, l'underfitting (ou sous-ajustement ) se produit lorsqu'un modèle ne peut pas saisir correctement la structure sous-jacente des données.*

**Prise en compte éventuelle des éléments suivants :**

- Nombre d'itérations dans le modèle
- La convergence
- Méthodes de sélection des variables de régression

5. **Séries temporelles ou chronologiques (time series)**

Une série temporelle est une suite de valeurs numériques représentant l'évolution d'une variable dans le temps. L'axe des abscisses est donc le temps. En ordonnée on a la valeur de la variable.

La série contient généralement une tendance, des variations saisonnières et des variation accidentelles.

A partir d'une série de données, on peut prédire la valeur de la variable dans le futur.

# Enoncé

Choisir une des méthodes citées ci-dessus.

Choisir un fichier de données sur lequel sera faite l'étude

Ecrire un programme en langage Python qui met en œuvre la méthode :

- Choisir un fichier de données sur lequel la méthode sera testée
- Lecture du fichier des données
- Dans le cas de le régression, le programme permet de saisir les données d'un nouvel exemple pour prédire son appartenance.

**Livrables :**

1 – Document de quelques pages expliquant le travail réalisé

2 – Le fichier des données

3 – le script Python commenté

4 – présentation orale (10 mn)

Travail à réaliser en binôme et à rendre au plus tard le 15 janvier.