

# Régression Linéaire Simple

La prédiction numérique est semblable à la classification.

- On construit un modèle dans une première phase.
- On utilise ce modèle pour prédire une valeur continue pour une instance donnée.

La classification réfère à la prédiction d'une classe catégorique

La prédiction numérique modèle des fonctions à valeur continue

L'instance contient les valeurs d'attributs connues.

Certains classifieurs peuvent être adaptés pour la prédiction numérique.

Exemple : KNN K-nearest neighbor

Parmi les méthodes d'analyses de régression :

Régression linéaire simple et multiple

Régression non linéaire

Modèle linéaire généralisé : Régression de Poisson et logistique

Régression Linéaire Simple :

Implique une variable de sortie Y, et un seul attribut X (variable d'entrée)

Le modèle a la forme linéaire :

$$Y = w_0 + w_1 X$$

$w_0$  et  $w_1$  sont les coefficients de la régression

Méthode des moindres carrés : On estime la meilleure ligne droite comme étant celle qui minimise l'erreur entre les données réelles et l'estimation de la ligne.

$$w_1 = \frac{\sum_{i=1}^{|D|} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|} (x_i - \bar{x})^2} \quad w_0 = \bar{y} - w_1 \bar{x}$$

D : la base d'apprentissage

X : Valeurs de la variable attribut

Y : Valeurs de la variable réponse

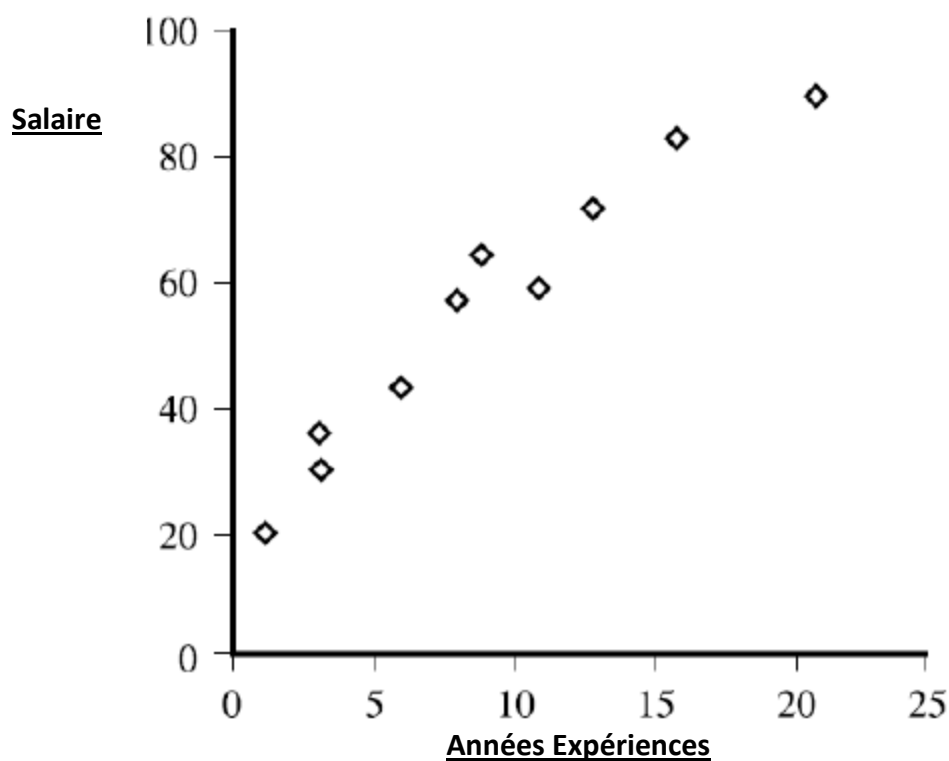
IDI : les points donnés de la forme  $(x_1, y_1), (x_2, y_2), \dots, (x_{IDI}, y_{IDI})$ .

$\bar{x}$  : la moyenne des valeurs  $x_i$

$\bar{y}$  : la moyenne des valeurs  $y_i$

Exemple : Le tableau suivant montre des paires de données où X est le nombre d'années d'expérience d'un titulaire d'un Master et Y est le salaire correspondant de ce Master.

| X Années Expériences | Y Salaire (en 1000DA) |
|----------------------|-----------------------|
| 3                    | 30                    |
| 8                    | 57                    |
| 9                    | 64                    |
| 13                   | 72                    |
| 3                    | 36                    |
| 6                    | 43                    |
| 11                   | 59                    |
| 21                   | 90                    |
| 1                    | 20                    |
| 16                   | 83                    |



Avec ces données, nous calculons :

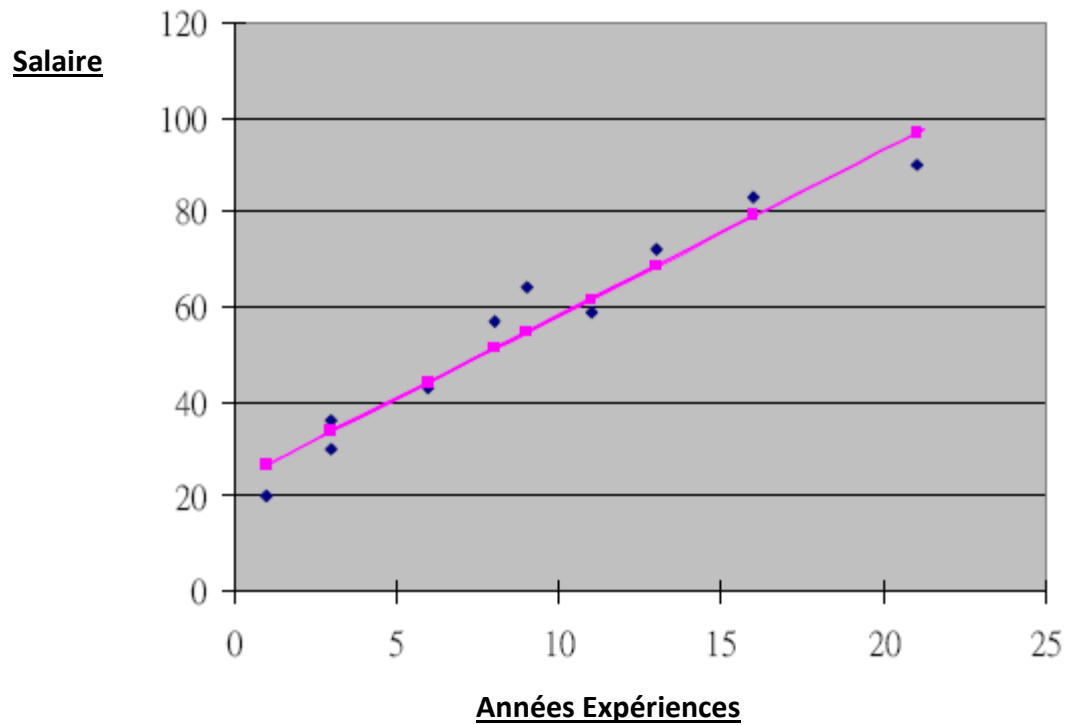
$$\bar{x} = 9.1 \text{ et } \bar{y} = 55.4$$

Nous avons donc :

$$w_1 = \frac{(3 - 9.1)(30 - 55.4) + (8 - 9.1)(57 - 55.4) + \dots + (16 - 9.1)(83 - 55.4)}{(3 - 9.1)^2 + (8 - 9.1)^2 + \dots + (16 - 9.1)^2} = 3.5$$

$$w_0 = 55.4 - (3.5)(9.1) = 23.6$$

$$Y = 23.6 + 3.5X$$



L'objectif de l'algorithme Régression Linéaire Simple était justement de trouver cette ligne (Modèle) qui va l'aider à prédire le salaire (classe numérique) de toute variable en entrée (années d'expériences).