FACULTÉ DES SCIENCES ET TECHNIQUES
DE MOHAMMEDIA
UNIVERSITÉ HASSAN II DE CASABLANCA

Régression pour l'apprentissage automatique

B. HILALI - brahill79@gmail.com / T. SABRI – sabritarik@gmail.com
Filière: Ingénierie Logicielle et Intégration des Systèmes Informatiques (ILISI) 24-25

Définition de la Régression

- La régression est une méthode d'apprentissage supervisé utilisée pour prédire une valeur continue (variable cible) à partir d'une ou plusieurs variables explicatives (features). Elle est particulièrement utile dans les cas où la relation entre les variables explicatives et la variable cible peut être modélisée mathématiquement.

- En d'autres termes, la régression cherche à apprendre une fonction f(X) qui associe un ensemble de variables d'entrée X={x1,x2,...,xn} à une variable de sortie y. Cette fonction est ensuite utilisée pour faire des prédictions sur de nouvelles données.

Caractéristiques Fondamentales de la Régression

- Variable Cible Continue
  - ✓prédire le prix d'une maison en fonction de ses caractéristiques (superficie, emplacement, nombre de chambres, etc.)
- Apprentissage Supervisé
  - ✓Les algorithmes de régression nécessitent un jeu de données étiquetées, c'est-à-dire que chaque observation doit inclure à la fois les variables explicatives et la valeur réelle de la variable cible.
- Modélisation Mathématique
  - ✓Existence d’une relation sous-jacente entre les variables explicatives et la variable cible.

    relation linéaire → régression linéaire

    relation non-linéaire → régression polynômiale, …

Caractéristiques Fondamentales de la Régression

- Évaluation des Performances

  Les performances des modèles de régression sont généralement évaluées à l'aide de métriques comme :

  ✓Erreur Quadratique Moyenne (MSE) :

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  ✓Coefficient de Détermination (R2) :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Types d'algorithmes de Régression

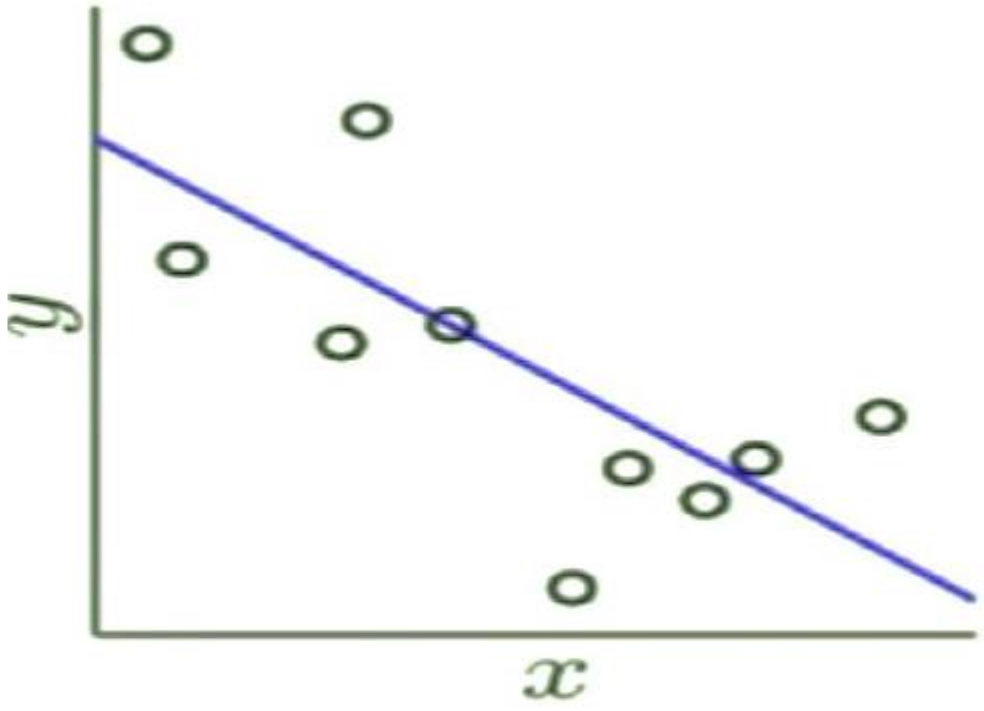
Régression Linéaire Simple

- Utilisée lorsque la relation entre une seule variable explicative (X) et la variable cible (y) est supposée linéaire.

- Formule :

$$y = \beta_0 + \beta_1 X + \epsilon$$

  - β0: Intercept (ordonnée à l'origine)
  - β1:Pente
  - ε: Terme d'erreur



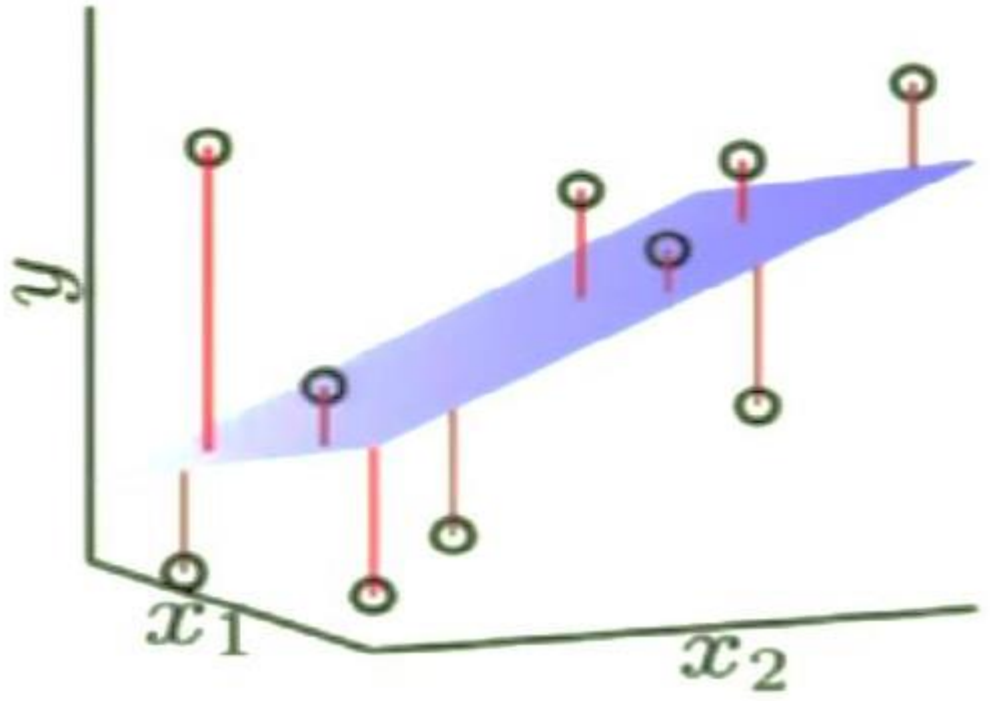
(a) une dimension (ligne)

Types d'algorithmes de Régression

Régression Linéaire Multiple

- Étendue de la régression linéaire simple, elle implique plusieurs variables explicatives (X1,X2,...,Xn).
- Formule :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon$$



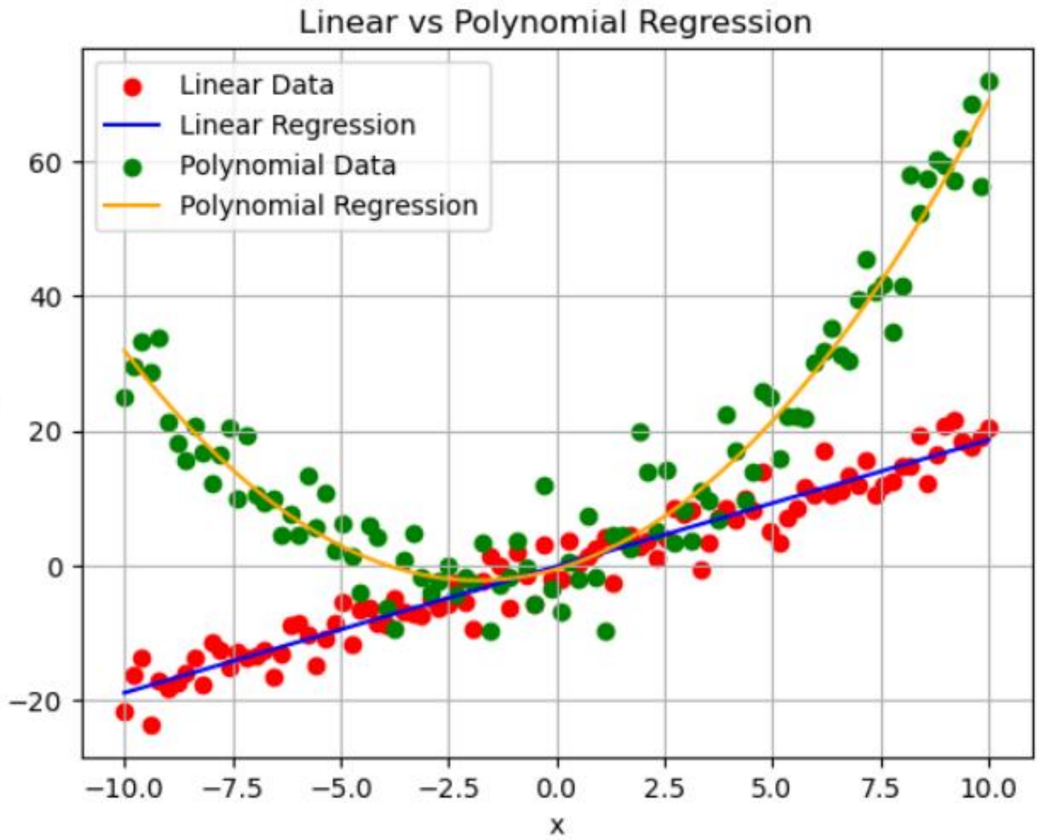
(b) deux dimensions (hyperplan)

Types d'algorithmes de Régression

Régression Polynomiale

- Utilisée lorsque la relation entre les variables explicatives et la variable cible n'est pas linéaire mais peut être approximée par un polynôme.
- Formule :

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \ldots + \beta_n X^n + \epsilon$$



Hypothèses de la Régression Linéaire

Pour que la régression linéaire fonctionne correctement, certaines hypothèses doivent être respectées :

- ✓Linéarité : La relation entre les variables explicatives et la variable cible doit être approximativement linéaire.
- ✓Indépendance des erreurs : Les résidus (erreurs) ne doivent pas être corrélés entre eux.
- ✓Homoscédasticité : La variance des erreurs doit être constante pour toutes les valeurs de X.
- ✓Normalité des erreurs : Les erreurs doivent suivre une distribution normale (surtout important pour les tests statistiques).
- ✓Absence de multicolinéarité : Les variables explicatives ne doivent pas être fortement corrélées entre elles.

Régression Linéaire simple

Utilisée lorsque la relation entre une seule variable explicative (X) et la variable cible (y) est supposée linéaire.

- Formule : $y=\beta 0+\beta 1X+\epsilon$
- $\beta 0$: Intercept (ordonnée à l'origine)
- $\beta 1$: Pente
- ϵ : Terme d'erreur

Régression Linéaire simple

Méthode d'Optimisation

Pour minimiser la fonction de coût, deux approches principales sont utilisées :

✓Méthode des Moindres Carrés Ordinaires

✓Descente de Gradient

Régression Linéaire

Méthode des Moindres Carrés Ordinaires

La méthode des moindres carrés ordinaires est une solution analytique pour trouver directement les coefficients optimaux. Elle repose sur la formule suivante :

$$\beta = (X^T X)^{-1} X^T y$$

- X : Matrice des variables explicatives (incluant une colonne de 1 pour l'intercept).
- y : Vecteur des valeurs réelles.
- β : Vecteur des coefficients.

Régression Linéaire

Méthode de la Descente de Gradient

La descente de gradient est une méthode itérative qui ajuste progressivement les coefficients pour minimiser la fonction de coût (l’Erreur Quadratique Moyenne (MSE)).

$$J(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Où :

n : Nombre d'observations.

yi : Valeur réelle pour l'observation i.

y^i= β0+β1X : Valeur prédite.

Régression Linéaire

Méthode de la Descente de Gradient

La descente de gradient est une méthode itérative qui ajuste progressivement les coefficients pour minimiser la fonction de coût. Les étapes sont les suivantes :

✓Initialiser les coefficients (β) avec des valeurs aléatoires.

✓Calculer le gradient de la fonction de coût par rapport à chaque coefficient :

$$\frac{\partial J(\beta)}{\partial \beta_j} = -\frac{2}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)X_{ij}$$

✓Mettre à jour les coefficients :

$$\beta_j := \beta_j - \alpha \frac{\partial J(\beta)}{\partial \beta_j}$$

Où α est le taux d'apprentissage.

✓Répéter jusqu'à convergence (lorsque les changements dans β deviennent négligeables).

Régression Linéaire

Évaluation du Modèle

Une fois les coefficients estimés, le modèle est évalué en utilisant des métriques telles que :

✓Erreur Quadratique Moyenne (MSE) : Mesure l'erreur moyenne entre les prédictions et les valeurs réelles.

✓Coefficient de Détermination (R2) : Indique la proportion de la variance expliquée par le modèle.