**Régression Linéaire : Définition et Fonctionnement**

La régression linéaire modélise la relation entre une variable dépendante yyy et une ou plusieurs variables indépendantes xxx en ajustant une ligne droite (également appelée "ligne de régression") à travers les points de données. L'équation de cette ligne est généralement de la forme :

Une image contenant texte, Police, ligne, capture d’écran

Description générée automatiquement

**Exemple en Python avec scikit-learn (Voir jupiter Notebook)**

Scikit-learn est une bibliothèque populaire pour le machine learning en Python. Elle offre des outils simples et efficaces pour l'analyse de données et la modélisation statistique, y compris la régression linéaire.

Imaginons un cas simple où nous avons des données sur la durée de vie restante de moteurs d'hélicoptères en fonction de l'heure totale de fonctionnement. Nous pouvons utiliser ces données pour entraîner un modèle de régression linéaire. Voici comment on pourrait faire cela en Python :

Dans le contexte de la régression linéaire, les matrices colonne (ou vecteurs colonne) sont souvent utilisées pour représenter les données d'entrée (features) et les étiquettes (labels) en raison de la manière dont les algorithmes de régression et d'apprentissage machine traitent les données. Voici pourquoi :

1. **Compatibilité avec les algorithmes** :
   * Les algorithmes de régression linéaire, tels que celui fourni par la bibliothèque scikit-learn, s'attendent à ce que les données d'entrée soient sous forme de matrice 2D où chaque colonne représente une caractéristique (feature) et chaque ligne représente une observation.
   * En utilisant une matrice colonne pour XXX, nous explicitons que chaque observation (heure de fonctionnement du moteur) est une caractéristique unique, ce qui est nécessaire pour l'algorithme de régression linéaire.
2. **Forme des données** :
   * Les algorithmes de machine learning nécessitent souvent une distinction claire entre les observations (lignes) et les caractéristiques (colonnes). Même si dans ce cas spécifique, chaque observation n'a qu'une seule caractéristique, il est toujours nécessaire de le formater en une matrice 2D.



1. **Compatibilité avec l'apprentissage supervisé** :
   * La régression linéaire implique la multiplication matricielle entre les coefficients (poids) et les caractéristiques des données d'entrée. Pour que cette multiplication soit valide, XXX doit être une matrice 2D.
2. **Consistance des prédictions** :
   * Lorsque vous effectuez des prédictions, la méthode predict s'attend également à ce que les nouvelles données soient sous forme de matrice 2D. C'est pourquoi np.array([[700]]) est utilisé pour les nouvelles données de prédiction, garantissant que les dimensions sont compatibles avec les attentes du modèle.

--