

# Régression linéaire

## fiche d'aide

Arts et Metiers



### A quoi sert la régression linéaire ?

La régression linéaire est une méthode statistique qui vise à prédire la valeur d’une variable expliquée en fonction de variables explicatives. Les avantages de la régression linéaire sont nombreux. Lorsque l’algorithme est entraîné, la prédiction est très rapide. Cependant, la complexité de l’algorithme augmente fortement avec le nombre de variables. De plus, contrairement à une régression polynomiale, le jeu de données doit suivre une relation linéaire pour obtenir une prédiction correcte.

### Comment fonctionne la régression linéaire ?

Pour effectuer une régression linéaire on peut chercher à minimiser l’erreur quadratique moyenne du jeu de test. l’erreur quadratique moyenne est définie par la formule suivante :

$$EQM = \sum_{i=0}^n \left( \frac{y_i - m}{\sigma_i} \right)^2$$

On en deduit les coefficients de la droite de régression linéaire et on obtient une droite de la forme :  $y = \alpha x + \beta$

### Avantages et Inconvénients

Un modele de régression lineaire presente selon les cas des avantages et des inconvenients.

#### Avantages

1. Simplicité: La régression linéaire est un modèle simple et facile à comprendre. Son interprétation est souvent intuitive, ce qui le rend accessible même pour ceux qui ne sont pas experts en statistiques ou en machine learning.
2. interprétabilité: Les coefficients de la régression linéaire représentent la force et la direction de la relation entre les variables, ce qui facilite l’interprétation des résultats.
3. Efficacité: Lorsque la relation entre les variables est linéaire, la régression linéaire peut fournir des résultats précis et robustes.

#### Inconvénients

1. Sensibilité: La régression linéaire peut être sensible aux valeurs aberrantes, ce qui signifie que des points de données extrêmes peuvent influencer considérablement les résultats.
2. Linéarité: Comme son nom l’indique, la régression linéaire suppose une relation linéaire entre les variables. Si cette hypothèse n’est pas respectée, les résultats peuvent être biaisés.
3. Limitation à des relations simples: La régression linéaire ne peut modéliser que des relations linéaires, ce qui la rend limitée pour des situations où les relations entre les variables sont non linéaires ou complexes.
4. Multicollinéarité: Lorsque les variables indépendantes sont fortement corrélées entre elles, cela peut entraîner des problèmes de multicollinéarité, affectant la stabilité des coefficients estimés.

### Exemple

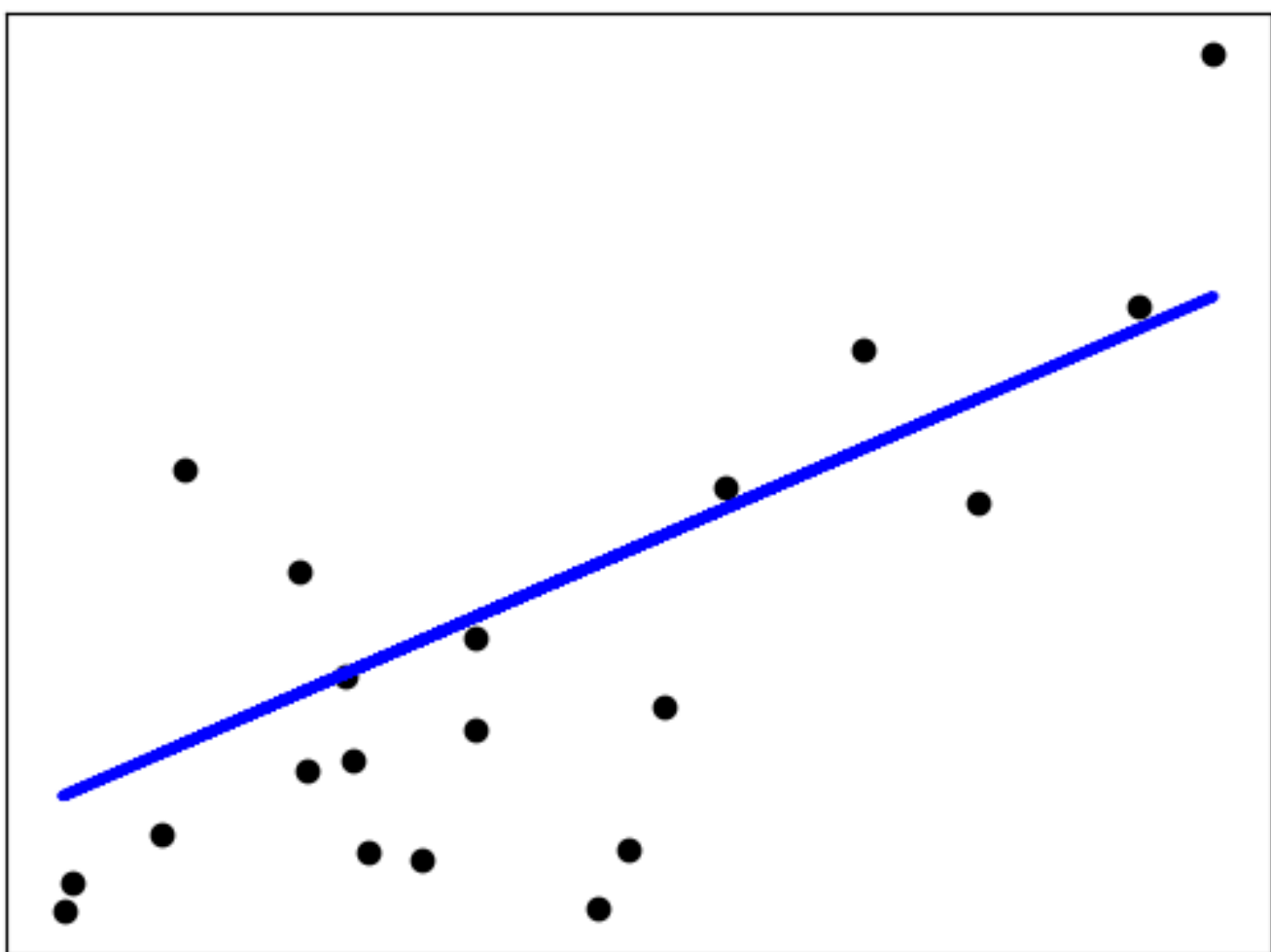


Figure 1: Exemple de régression linéaire avec sklearn

On peut mettre en place le pseudo-code suivant:

### References

- [1]
- [2]
- [3]
- [4]
- [5]
- [6]
- [7]
- [8]
- [9]
- [10]
- [11]
- [12]
- [13]

### Pour aller plus loin

1. Scikit-learn.
2. Wikipedia (la page française n’est pas très bonne)