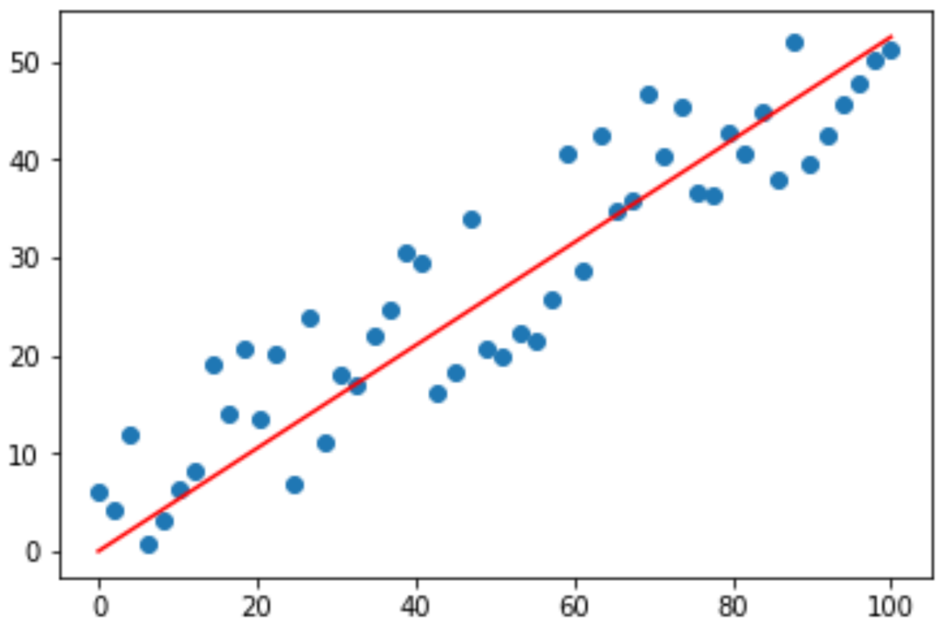
La Régression Linéaire

Travail réalisé par

Adam EL MOUEDDEN

27 octobre 2025



Projet de Mathématiques Appliquées

Analyse des méthodes de régression linéaire et applications pratiques

Table des matières

1 Introduction

Ce projet vise à approfondir la compréhension de la régression linéaire, l'une des méthodes statistiques les plus fondamentales et les plus utilisées en science des données. La régression linéaire constitue la pierre angulaire de nombreux modèles prédictifs et trouve des applications dans pratiquement tous les domaines scientifiques.

L'objectif principal de ce travail est double :

— Comprendre les fondements mathématiques de la régression linéaire

— Citer l'unes de ses applications pratiques dans divers domaines scientifiques

2 Régression Linéaire Simple

2.1 Le but

On dispose d'un ensemble de n observations :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

où :

— x_i est la variable explicative (ou indépendante),

— y_i est la variable à expliquer (ou dépendante).

On suppose que la relation entre x et y est approximativement linéaire, c'est-à-dire :

$$y_i \approx ax_i + b$$

où :

— a = pente (slope),

— b = ordonnée à l'origine (intercept).

Notre objectif est de trouver les valeurs de a et b qui « collent » le mieux aux données.

2.2 Modèle mathématique

On modélise :

$$y_i = ax_i + b + \varepsilon_i$$

avec :

— ε_i : l'erreur (ou résidu), supposée centrée (moyenne nulle).

$$\text{(forme matricielle : } \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \text{a.} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \text{)}$$

2.3 Méthode des moindres carrés (Least Squares)

On cherche à minimiser la somme des carrés des erreurs :

$$S(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2 = \sum_{i=1}^{n} \epsilon_i^2$$

On veut :

$$\min_{a,b} S(a,b) \quad (= \min_{a,b} ||\epsilon||_2^2 \quad avec \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}))$$

2.4 Calcul analytique

On dérive $S(a,b)$ par rapport à a et b et on annule les dérivées :

$$\frac{\partial S}{\partial a} = -2 \sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0$$
$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

En simplifiant ces deux équations :

$$\sum y_i = a \sum x_i + nb$$
$$\sum x_i y_i = a \sum x_i^2 + b \sum x_i$$

C'est un système linéaire à deux inconnues a, b.

2.5 Résolution du système

En notant :

$$\bar{x} = \frac{1}{n} \sum x_i, \quad \bar{y} = \frac{1}{n} \sum y_i$$

On obtient les formules classiques :

$$a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$
$$b = \bar{y} - a\bar{x}$$

2.6 Interprétation géométrique

La droite :

$$y = ax + b$$

est la droite qui minimise la distance verticale entre les points observés et la droite. Graphiquement, c'est la droite « la plus proche » des points selon le critère des moindres carrés.

2.7 Qualité de l'ajustement

On mesure la qualité de la régression par le coefficient de détermination R^2 :

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

où $\hat{y}_i = ax_i + b$.

- Si $R^2 = 1 \rightarrow$ ajustement parfait.
- Si $R^2 = 0 \rightarrow$ la droite ne prédit rien (moyenne constante).