Méthode des moindres carrés
Présentation, démonstration et forme matricielle

1 Introduction

La méthode des moindres carrés est une technique mathématique permettant d'estimer les paramètres d'un modèle en minimisant l'erreur entre les valeurs observées et les valeurs prédites par le modèle.

2 Problème fondamental

On considère un ensemble de données expérimentales :

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

On cherche à approximer ces données par une fonction linéaire :

$$y = ax + b$$

En général, il n'existe pas de droite passant exactement par tous les points. On cherche donc la droite qui **approxime au mieux** les données.

3 Principe de la méthode des moindres carrés

Pour chaque point i, on définit le résidu :

$$e_i = y_i - (ax_i + b)$$

La méthode des moindres carrés consiste à minimiser :

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

4 Démonstration (cas scalaire)

On cherche :

$$\frac{\partial S}{\partial a} = 0 \quad \text{et} \quad \frac{\partial S}{\partial b} = 0$$

4.1 Dérivée par rapport à a

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

Condition d'optimalité :

$$\sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0$$

4.2 Dérivée par rapport à b

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

Condition d'optimalité :

$$\sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

5 Équations normales

On obtient le système :

$$\begin{cases} a \sum x_i^2 + b \sum x_i = \sum x_i y_i \\ a \sum x_i + nb = \sum y_i \end{cases}$$

6 Solution explicite

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$b = \bar{y} - a\bar{x}$$

avec :

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

7 Forme matricielle du problème

7.1 Modèle

$$Y = X\beta + \varepsilon$$

où :
- $Y \in \mathbb{R}^n$ est le vecteur des observations
- $X \in \mathbb{R}^{n \times p}$ est la matrice des variables explicatives
- $\beta \in \mathbb{R}^p$ est le vecteur des paramètres
- ε est le bruit

8 Fonction coût matricielle

$$J(\beta) = \|Y - X\beta\|^2$$

Soit :

$$J(\beta) = (Y - X\beta)^T (Y - X\beta)$$

9 Démonstration matricielle

$$J(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta$$

10 Calcul du gradient

On utilise les identités :
— $\nabla_\beta(\beta^T a) = a$
— $\nabla_\beta(\beta^T A\beta) = 2A\beta$ si A est symétrique
Comme $X^T X$ est symétrique :

$$\nabla_\beta J(\beta) = -2X^T Y + 2X^T X\beta$$

Condition d'optimalité :

$$\nabla_\beta J(\beta) = 0$$

Donc :

$$X^T X\beta = X^T Y$$

11 Solution des moindres carrés

Si $X^T X$ est inversible :

$$\boxed{\hat{\beta} = (X^T X)^{-1} X^T Y}$$

12 Interprétation géométrique

— $X\hat{\beta}$ est la projection orthogonale de Y sur l'espace engendré par les colonnes de X
— Le résidu $\hat{r} = Y - X\hat{\beta}$ est orthogonal à cet espace :

$$X^T(Y - X\hat{\beta}) = 0$$

13 Cas général : matrice non inversible

Si $X^T X$ n'est pas inversible :
— les colonnes de X sont linéairement dépendantes
— il existe une infinité de solutions
On utilise alors la **pseudo-inverse de Moore–Penrose** :

$$\hat{\beta} = X^+ Y$$

Calculée via la décomposition en valeurs singulières (SVD).

14 Conclusion

La méthode des moindres carrés :
— fournit une solution optimale au sens quadratique
— possède une interprétation géométrique claire
— est la base de la régression linéaire moderne
Elle s'étend naturellement aux modèles non linéaires et probabilistes.