

Série de Td N°3

Data Mining – Régression linéaire – 2022/2023 – R.M : Ali LALOUCI

Exercice 1 (calcul matriciel)

1) Soit $A = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$ et $B = \begin{pmatrix} 0 & 2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$

- a) Calculer AB puis $(AB)^t$
- b) Calculer $A^t B^t$ et vérifier que $AB^T = A^t B^t$

2) Soient $A = \begin{pmatrix} -2 & 4 \\ 3 & -6 \end{pmatrix}$ et $B = \begin{pmatrix} 2 & 2 \\ 1 & 1 \end{pmatrix}$

- a) Vérifier que $AB = 0$ bien que $A \neq 0$ et $B \neq 0$
 - b) Calculer BA et vérifier que $AB \neq BA$
- 3) Calculer les inverses des matrices suivantes :

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 2 & 3 & 2 \end{pmatrix}$$

Exercice 2 (Régression linéaire simple)

Une étude de la température de fonctionnement d'un procédé chimique sur le rendement du produit a donné les valeurs suivantes pour la température et le rendement correspondant :

Température °C	100	110	120	130	140	150	160	170	180	190
Rendement %	45	51	54	61	66	70	74	78	85	89

Questions :

- 1) Donner une représentation graphique de ces données.
- 2) Trouver la fonction de régression linéaire par la méthode des moindres carrés, qui permet d'associer à la température la valeur de rendement correspondante.
- 3) Utilisant cette fonction de régression, prédire la valeur de rendement pour la température 80°C.
- 4) Déterminer (en utilisant la droite de régression) quand la valeur de rendement sera supérieure à 100.

Exercice 3 (Régression linéaire multiple)

On souhaite expliquer une variable dépendante Y (par exemple : consommation d'énergie, note d'examen, prix d'un appartement...) à l'aide de deux variables explicatives X_1 et X_2 . On donne les données suivantes pour un échantillon de 6 observations.

$$Y = \begin{pmatrix} 12 \\ 15 \\ 17 \\ 18 \\ 22 \\ 24 \end{pmatrix} \quad X = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 1 & 4 \\ 1 & 3 & 2 \\ 1 & 4 & 5 \\ 1 & 5 & 3 \\ 1 & 6 & 4 \end{pmatrix}$$

- La première colonne est la constante (terme indépendant).
- La seconde colonne correspond à la variable X_1 .
- La troisième colonne correspond à la variable X_2 .

Questions :

- 1) Estimer le vecteur des coefficients $\hat{\beta}$.
- 2) Écrire l'équation estimée du modèle de régression
- 3) Calculer les valeurs ajustées $\hat{Y}$