

Devoir maison n°12

A rendre le 31/01/2023

Exercice 1

Dans cet exercice, se place dans \mathbb{R}^n munit du produit scalaire canonique.

On considère n points de \mathbb{R}^2 , $M_i : \begin{pmatrix} x_i \\ y_i \end{pmatrix}$ pour $i \in \llbracket 1, n \rrbracket$, et on cherche une droite $\mathcal{D} : y = ax + b$ qui serait une “meilleure approximation affine” du nuage de point $\{M_i, i \in \llbracket 1, n \rrbracket\}$. Pour cela on cherche à minimiser la quantité (dépendant de a et b).

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

qui représente la somme des carrés des distances obtenues par projection verticale. C’est la méthode des moindres carrés.

On considère $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ et $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$.

1. Pour $a, b \in \mathbb{R}$ fixés, exprimer $S(a, b)$ sous forme d’une norme faisant intervenir X, Y, U .
2. On note $F = \text{Vect}(U, X)$. A quelle condition sur les points M_i , l’espace F est-il un plan de \mathbb{R}^n ?
3. On note $u_{a,b} = aX + bU$. On suppose que $Y - u_{a,b} \perp F$. Montrer que si $v \in F$, alors $\|Y - v\| \geq \|Y - u_{a,b}\|$ en utilisant le théorème de Pythagore.
4. On cherche donc (a, b) tels que $Y - u_{a,b} \perp F$. Montrer que le couple (a, b) cherché est solution du système

$$\begin{cases} a \sum_{i=1}^n x_i + nb = \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i \end{cases}$$

5. Montrer que si les M_i n’ont pas tous la même abscisse, alors le système de la question précédente (d’inconnues a, b) est un système de Cramer.¹
6. En notant M la matrice du système précédent, exprimer l’unique solution du système en fonction de M et de produits scalaires.
7. Écrire en python une fonction `regression(X, Y)` qui prend comme arguments une liste d’abscisses X (pas toutes égales) et une liste d’ordonnées Y de même longueur, et retournant les flottants a, b correspondants aux points dont les coordonnées sont données.

On pourra utiliser `numpy.linalg.solve` pour résoudre le système considéré.

8. Question bonus : tracer sur un même graphique le nuage de point et la droite calculée.

1. On pourra aller réviser l’inégalité de Cauchy-Schwartz, ainsi que son cas d’égalité.

Indications

1. Écrire un vecteur dont $S(a, b)$ est la norme puis exprimer ce vecteur en fonction des données de l'énoncé.
2. On pourra se servir du fait que $U \neq 0$.
3. Écrire $Y - v$ en faisant intervenir $u_{a,b}$ et on utilisera l'hypothèse.
4. $w \perp F$ ssi
- 5.
6. Il s'agit d'une question de révision concernant l'écriture matricielle d'un système.
7. On pourra utiliser la fonction `sum` et des listes par compréhension. Voir l'aide de la fonction proposée pour plus de détails sur son fonctionnement.