

# exam\_blanc

April 3, 2024

## 1) Clustering

Soient les données

$$\begin{pmatrix} x & y \\ 1 & 1 \\ 3 & 6 \\ 5 & 6 \\ 2 & 1 \\ 5 & 5 \\ 1 & 2 \\ 7 & 5 \end{pmatrix}$$

Appliquez l'algorithme des k-means pour  $K = 3$  d'abord avec ces centres :

$$\begin{pmatrix} x & y \\ 8 & 4 \\ 6 & 5 \\ 1 & 4 \end{pmatrix}$$

Puis avec ceux-ci :

$$\begin{pmatrix} x & y \\ 4 & 9 \\ 2 & 3 \\ 2 & 8 \end{pmatrix}$$

Commentez.

Comparez ensuite avec le clustering hiérarchique "complete linkage".

## 0.1 2) Chaînes de Markov

Une chanteuse d'opéra doit donner une longue série de concerts. Son tempérament d'artiste la pousse à vouloir tous les soirs arrêter les concerts, et ce avec une probabilité de  $1/2$ . Une fois qu'elle a décidé d'arrêter, elle ne chantera pas de nouveau jusqu'à ce que l'organisateur la convainque de son admiration. Pour cela il lui envoie des fleurs chaque jour jusqu'à ce qu'elle revienne. Des fleurs coutant  $x$  milliers d'euros,  $0 \leq x \leq 1$ , amènent une réconciliation avec la probabilité  $p = \sqrt{x}$ . L'organisateur fait 750 euros de bénéfice à chaque représentation donnée.

Combien doit-il dépenser en fleurs ?

(Source : [https://philcarmona.github.io/extra/TouslesexosetdevoirsX22M010\\_2018.pdf](https://philcarmona.github.io/extra/TouslesexosetdevoirsX22M010_2018.pdf) )

### 0.2 3) k plus proches voisins

Soient les données

$$\begin{pmatrix} x & y & \text{classe} \\ 3 & 0 & 1 \\ 4 & 3 & 2 \\ 5 & 4 & 2 \\ 9 & 6 & 2 \\ 1 & 7 & 1 \\ 6 & 1 & 1 \\ 0 & 5 & 1 \\ 7 & 2 & 2 \\ 8 & 9 & 1 \\ 2 & 8 & 2 \end{pmatrix}$$

Prenant  $k = 3$ , considérez d'abord les 4 premières rangées comme l'ensemble de test, puis les 4 dernières. Dans les deux cas calculez le taux d'erreur. Commentaires ?

### 0.3 4) Arbre de régression

Soient les données

$$\begin{pmatrix} x & y & z \text{ (cible)} \\ 3 & A & 1 \\ 4 & A & 2 \\ 5 & B & 3 \\ 9 & B & 4 \\ 1 & A & 5 \\ 0 & C & 6 \\ 7 & C & 7 \\ 8 & C & 8 \\ 2 & B & 9 \end{pmatrix}$$

Construisez l'arbre de régression associé.