

- | |
|--|
| <ul style="list-style-type: none">- Durée : 3h- Calculatrice autorisée- Documents autorisés :<ul style="list-style-type: none">- supports de cours et notes personnelles- votre voisin n'est pas un document- Barème indicatif sur 30 points |
|--|

1 Clustering (13 points)

X	1	2	9	12	20
---	---	---	---	----	----

1. (7 points) K-Means

(a) Appliquez l'algorithme des K-means avec les valeurs de k et les points de départ suivants :

- i. $k = 2, \mu_1 = 1, \mu_2 = 20$.
- ii. $k = 3, \mu_1 = 1, \mu_2 = 12, \mu_3 = 20$.
- iii. $k = 4, \mu_1 = 1, \mu_2 = 9, \mu_3 = 12, \mu_4 = 20$.

(b) On aimerait maintenant comparer la qualité de ces regroupements. Pour cela, on recommence par regarder l'inertie intra-cluster.

- i. Calculer cette valeur pour les 3 regroupements précédents.
- ii. En utilisant ce critère, quel serait le meilleur regroupement possible ? est-ce que cela vous paraît réaliste ?

(c) S'inspirant du critère BIC, quelqu'un propose de rajouter le terme suivant au critère précédent : $+2kN \log N$ (où N est le nombre de données).

- i. Expliquer l'utilité de ce terme
- ii. Calculer la valeur du nouveau critère pour vos 3 regroupements. Qu'en concluez-vous

2. (6 points) Classification Hiérarchique

(a) Classification Hiérarchique Ascendante

- i. Appliquer l'algorithme de classification hiérarchique ascendante en utilisant le saut minimal et tracer le dendrogramme correspondant.
- ii. Idem avec le saut maximal.

(b) Classification Hiérarchique Descendante

Soit un algorithme de classification hiérarchique descendante qui recherche à chaque itération la meilleure façon de couper un ensemble de points en deux parties

- i. Détailler la première itération de cet algorithme (en utilisant un saut minimal)
- ii. Expliquer l'utilité de cet algorithme.