Une image contenant texte

Description générée automatiquement

**Nom :**

# Master 1 SD Prénom :

**Analyse et visualisation de données**

Séance de TP 2

K-means, K-médoïdes, critères de qualité

1. **Programmer la méthode des k-moyennes**

Rappeler l’algorithme et décrire les choix d’implémentation en termes de structure et types de données manipulées et passés en argument d’entrées et en valeur de retour de méthode.

Donner votre programme en annexe.

1. **Examiner la stabilité de l’algorithme sur plusieurs exécutions**

Pour comparer les exécutions pour une même valeur de K, on s’aidera des critères suivants :

Nombre d’itérations jusqu’à converge pour un seuil de 0,001

Pourcentage de variance expliquée

Position des centroïdes

- Effectuer cette analyse pour le dataset IRIS et pour le dataset Breast Cancer

1. Initialisation k-means++

Rappeler le principe de la méthode d’initialisation k-means++

Décrire vos choix d’implémentation en termes de structure et type de données manipulées et passés en arguments d’entrée et en valeurs de retour de méthode

Donner votre programme en annexe.

1. **Examiner la stabilité de l’algorithme avec kmeans++**

Pour comparer les exécutions pour une même valeur de K, on s’aidera des critères suivants :

Nombre d’itérations jusqu’à converge pour un seuil de 0,001

Pourcentage de variance expliquée

Position des centroïdes

- Effectuer cette analyse pour le dataset IRIS et pour le dataset Breast cancer

- Comparer ces résultats avec ceux obtenus à la question 2

1. **Programmer la méthode des k-médoïdes++**

Rappeler l’algorithme et décrire les choix d’implémentation en termes de structure et types de données manipulées et passés en argument d’entrées et en valeur de retour de méthode.

Donner votre programme en annexe.

1. **Examiner la stabilité de l’algorithme avec k-medoïdes+**

Pour comparer les exécutions pour une même valeur de K, on s’aidera des critères suivants

Nombre d’itérations jusqu’à converge pour un seuil de 0,001

Pourcentage de variance expliquée

Position des centroïdes

- Quel est l’apport de l’initialisation k-means++ sur les k-medoïdes ?

- Effectuer cette analyse pour le dataset IRIS et pour le dataset Breast cancer

1. **Vérifier le bon fonctionnement de vos programmes en les comparant aux méthodes de scikit learn**

Vous choisirez un exemple d’exécution et comparerez les résultats obtenus par votre méthode et la méthode équivalente de scikit-learn.