

Exercice 1

Un site de commerce en ligne souhaite analyser les critiques de produits de ses utilisateurs. Chaque critique est représentée par deux caractéristiques principales : la note (évaluée de 1 à 5) et le nombre de mots dans le texte de la critique. Les données correspondantes sont présentées dans le tableau suivant :

Note (1-5)	Nombre de mots
5	120
3	80
4	150
1	45
5	200
2	60
4	110
3	90
5	180
1	30

1. Choisissez initialement les deux premiers objets comme centroïdes initiaux. Exécutez l'algorithme k-means pour 2 itérations en utilisant la distance euclidienne. À la fin de chaque itération :
 - Donnez les nouveaux clusters.
 - Donnez les centres des nouveaux clusters.
2. Étant donné les paramètres finaux après les 2 itérations, à quel cluster appartiendrait le point (2,240) ?