

Corrigé type

Questions (10 points)

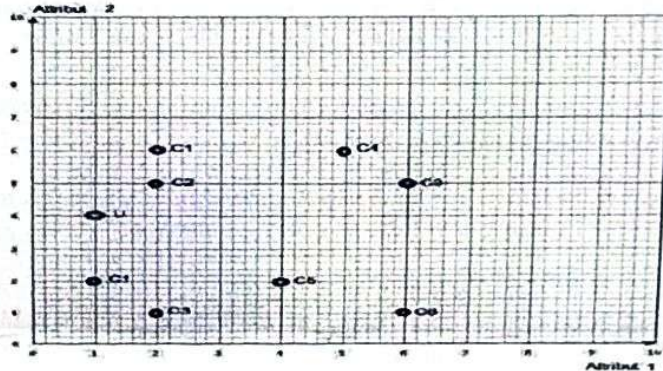
1. L'objectif de la classification est d'identifier les classes auxquelles appartiennent des objets à partir de traits descriptifs (attributs, caractéristiques, features). (2 points)
2. Dans les méthodes supervisées: les classes sont connues et l'on dispose d'exemples $\sqrt{a^2 + b^2}$ de chaque classe (fourni par un expert). Dans les méthodes non supervisées : les classes (clusters) ne sont pas définies à l'avance. On ne dispose pas d'exemples d'apprentissage. Les clusters sont construits en calculant la similarité entre les objets. (4 points)
3. (4 points)

Algorithme : K-means

1. Sélectionner K points comme centroïdes initiaux;
2. Former K clusters en assignant chaque point au centroïde le plus proche;
3. Recalculer le centroïde de chaque cluster nouvellement formé;
4. Répéter les étapes 2 et 3 jusqu'à ce qu'aucun centroïde ne change.

Exercice 1 (10 points)

1. (2 points)



3. (4 points)

On calcule la distance euclidienne qui sépare le point U à chacun des points :

Distance	Expression	Valeur
Distance(U, point 1)	$\text{Sqrt}((1-1)^2 + ((4-2)^2)$	02,00
Distance(U, point 2)	$\text{Sqrt}((2-1)^2 + ((6-2)^2)$	02,24
Distance(U, point 3)	$\text{Sqrt}((2-1)^2 + ((5-2)^2)$	01,41
Distance(U, point 4)	$\text{Sqrt}((2-1)^2 + ((1-2)^2)$	03,16
Distance(U, point 5)	$\text{Sqrt}((4-1)^2 + ((2-2)^2)$	03,61
Distance(U, point 6)	$\text{Sqrt}((5-1)^2 + ((6-2)^2)$	04,47
Distance(U, point 7)	$\text{Sqrt}((6-1)^2 + ((5-2)^2)$	05,10
Distance(U, point 8)	$\text{Sqrt}((6-1)^2 + ((4-2)^2)$	05,83

On garde les k ($k=3$) plus proches voisins (ceux qui ont la plus courte distance avec U). Il s'agit des points (1, 2 et 3). Parmi ces 3 points, c'est la classe C1 qui est majoritaire (2 voix pour C1, contre 1 voix pour C2). Donc le point U sera affecté à la classe C1.

4. (4 points)

Pour les k points voisins du point U trouvés à l'étape précédente, on calcule l'inverse de la distance au carré qui sépare le point U à chacun de ces points :

Distance	Distance d	$1/d^2$
Distance(U, point 1)	02,00	00,25
Distance(U, point 2)	02,24	00,20
Distance(U, point 3)	01,41	00,50

Pour la classe C1, la somme des poids pondérés est $0,25 + 0,20 = 0,45$. Pour la classe C2, la somme est 0,50. Le point U sera donc affecté à la classe 2.