

Data Mining

Exercice 1 : KNN

On souhaite utiliser l'algorithme K-NN pour prédire l'état d'un patient dépendamment de différents attributs. Le tableau ci-dessous comprend les informations concernant les trois attributs ainsi que l'état de 8 patients :

Travail à faire :

1. Classer les nouvelles observations **P9** et **P10** en appliquant l'algorithme K-NN. (Détaillez les calculs).
2. Utiliser la distance euclidienne qui a la formule suivante :

$$D_E = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$
3. Considérer la valeur : k=3 et par la suite k=5.

ID	Attribut 1	Attribut 2	Attribut 3	Etat
P1	3	1	1	Négative
P2	1	1	0	Négative
P3	6	7	1	Positive
P4	1	1	1	Négative
P5	10	3	3	Positive
P6	8	2	3	Positive
P7	6	8	4	Positive
P8	2	1	1	Négative
P9	1	1	1	?
P10	1	10	1	?

Exercice 2 : K-Means

Soit l'ensemble de données suivants constitué des deux attributs « Attribut 1 et Attribut 2 ». L'attribut ID ne va pas rentrer dans les calculs :

ID	Attribut 1	Attribut 2
A1	2	10
A2	2	5
A3	8	4
A4	5	8
A5	7	5
A6	6	4
A7	1	2
A8	4	9

On veut utiliser l'algorithme K-Means pour répartir les données en trois (3) clusters. On choisit d'une manière aléatoire les centres initiaux suivants des trois clusters respectivement :

M1 = (2, 10), M2 = (5, 8) et M3 = (1, 2)

La distance d entre les instances et les centres est calculée en utilisant la distance Manhattan : $D_E = \sum_{i=1}^n |x_i - y_i|$

Travail à faire :

1. Appliquer l'algorithme K-Means et montrez toutes les étapes de calcul.
2. Donnez le résultat final et précisez le nombre d'itérations qui ont été nécessaires pour arriver à une situation de stabilité.

Exercice 3 : Règles d'association

Soit le problème suivant :

Produits = {A, B, C, D, E, F}

Transactions = {AB, ABCD, ABD, ABDF, ACDE, BCDF}

Transactions	A	B	C	D	E	F
T1	1	1	0	0	0	0
T2	1	1	1	1	0	0
T3	1	1	0	1	0	0
T4	1	1	0	1	0	1
T5	1	0	1	1	1	0
T6	0	1	1	1	0	1

Avec : MINSUP = $\frac{1}{2}$, MINCONF = 75%

1. Donner l'ensemble des items et ItemSets fréquents.
2. Donner l'ensemble des règles valides.