

Modules "Datamining" " Fouille & Extraction de données " (M2-IL-ISIA-RFIA)

Fiche 6 : KNN (K nearest neighbors)

Exercice 1 : Rappelez l'objectif du modèle KNN.

Exercice 2 : Comment sont recherchés les K plus proches voisins d'une certaine donnée a ?.

Exercice 3 : Quelles sont les principales recommandations à faire sur le choix de la valeur k ?.

Exercice 4 : Le tableau suivant contient des données sur des individus d'une population décrits selon deux attributs : attribut 1 et attribut 2. La classe d'un individu peut être : C1, ou C2, ... ou C6.

Tableau des données :

N°	Attribut 1	Attribut 2	Classe
1	1	2	C1
2	2	6	C1
3	2	5	C2
4	2	1	C3

5	4	2	C5
6	5	6	C4
7	6	5	C3
8	6	1	C6

- Représentez sur le plan les données du tableau précédent.
- On veut classer un nouvel individu U ayant comme attributs (1, 4) en utilisant la méthode KNN. Quelle sera la classe de U si on choisit $k=3$? . Justifiez.
- On utilise maintenant la variante de KNN qui utilise la distance $1/d^2$ (inverse de la distance au carré) pour calculer les voisins. Quelle sera la classe de U avec $k=3$? . Justifiez .

Exercice 5 : Le tableau suivant contient des données sur 24 fermiers pour montrer s'ils peuvent acheter une machine agricole (tracteur). Chaque individu est représenté par son revenu (en millier de DA) et la superficie de sa ferme (en hectare). La classe 1 représente les fermiers qui peuvent être propriétaires d'un tracteur, et la classe 2 représente les fermiers qui ne peuvent pas l'être.

N°	Revenu	Superficie	Classe
01	60	18.4	C1
02	85.5	16.8	C1
03	64.8	21.6	C1
04	61.5	20.8	C1
05	87	23.6	C1
06	110.1	19.2	C1
07	108	17.6	C1
08	82.8	22.4	C1
09	69	20	C1
10	93	20.8	C1
11	51	22	C1
12	81	20	C1

13	75	19.6	C2
14	52.8	20.8	C2
15	64.8	17.2	C2
16	43.2	20.4	C2
17	84	17.6	C2
18	49.2	17.6	C2
19	59.4	16	C2
20	66	18.4	C2
21	47.4	16.4	C2
22	33	18.8	C2
23	51	14	C2
24	63	14.8	C2

- Représenter les données sur un plan.
- Utiliser l'algorithme KNN, avec $k=1$, pour connaître la classe d'un individu ayant un revenu=52.8 et une superficie=20.8 . Que remarque-t-on ?.
- On souhaite recalculer la classe de chacun des 24 individus en utilisant KNN , avec $k=1$. Calculer le nombre et le taux d'erreurs de classification faites ?.

Exercice 6 : Reprendre les données de l'exercice 5 sur Weka, avec le programme KNN et $k=1$. Quel est le nombre et le taux d'erreur de classification.

- Répéter l'expérience en faisant varier k , avec toutes les valeurs impaires allant de 1 à 25. Renseigner le tableau suivant :

K	Nombre d'erreurs de classification	% erreur de classification
1		
3		
...		
25		

- Discuter les résultats et déduire la meilleure valeur de k .