

Examen__AD__M2__MIAGE

27 mars 2023

1 Clustering

Données :

x1	x2
4	7
1	6
6	6
2	6
5	2
3	2

Appliquez en détaillant raisonnablement chaque étape l'algorithme des k-means avec $K = 2$. Prenez comme centres initiaux $c_1 = (1, 6)$ et $c_2 = (6, 6)$ dans un premier temps. Choisissez ensuite $c_1 = (4, 3)$ et $c_2 = (4, 7)$. Que remarquez-vous ?

Appliquez ensuite l'algorithme CAH "complete linkage" sur ces mêmes données. Dessinez le dendrogramme.

2 Chaînes de Markov

On considère un jeu de billard simplifié démarrant avec seulement 4 boules. Pour gagner Alice doit sortir les 2 boules vertes, et Bob les 2 boules bleues.

À chaque tour un joueur sort une seule de ses boules avec probabilité p , ou rien. Cette probabilité vaut $p_A = 0.4$ pour Alice et $p_B = 0.3$ pour Bob. Lorsqu'un joueur parvient à sortir une de ses boules, il rejoue.

- Modélisez la situation par une chaîne de Markov.
- Quels sont les états transitoires ? Absorbants ?
- Expliquez comment obtenir la probabilité qu'Alice gagne.
- Expliquez également comment déterminer la durée moyenne du jeu.

On considère à présent qu'un joueur peut aussi sortir une boule adverse pendant son tour.

- Modifiez le modèle pour tenir compte de cela (probabilités q_A, q_B).
- Comment est affectée (qualitativement) la durée du jeu ? Justifiez.

3 k-PPV en classification

On choisira $k = 3$ (trois voisins).

Données :

x1	x2	y
2	2	A
4	5	B
7	7	A
4	8	B
9	10	A
5	1	B
5	3	A
1	9	B
8	10	A
3	7	B
10	8	A
8	1	B

Considérant comme ensemble d'entraînement les lignes 1 à 8, calculez l'erreur E réalisée sur l'ensemble de test (lignes 9 à 12), c'est-à-dire le taux d'erreur : $\# \text{mal classés} / \# \text{total}$.

Répétez ensuite l'opération avec pour ensemble d'entraînement les lignes 5 à 12.

4 Arbres de décision

Données artificiellement générées avec A variable catégorielle et B variable réelle :

A	B	Cible
X	1	1
Y	6	1
Y	2	1
X	9	2
Z	5	2
Z	3	2
X	10	3
Z	7	3
Y	8	3

Construisez un arbre de décision selon la méthode vue en cours, en utilisant l'indice de Gini :

$$I = 1 - \sum_{k=1}^K p_k^2$$

où p_k est la proportion de la classe k dans l'ensemble considéré.