

TD n 4 : apprentissage artificiel

Exercice 1

Nous avons deux ensembles de points à 2 dimensions représentant deux classes différentes :

$$C_1 = \begin{pmatrix} 1 & 2 & 2 & 2 & 3 & 10 & 15 & 16 & 16 & 16 & 17 \\ 2 & 1 & 2 & 3 & 2 & 2 & 2 & 1 & 2 & 3 & 2 \end{pmatrix}$$

$$C_2 = \begin{pmatrix} 6 & 6 & 6 & 7 & 7 & 7 & 8 & 8 & 9 & 9 & 10 & 10 \\ 1 & 3 & 2 & 1 & 2 & 3 & 1 & 3 & 1 & 3 & 1 & 3 \end{pmatrix}$$

1. Calculer le vecteur moyen de classe (centroïde) pour chaque ensemble et classer le point (9.6, 2) en utilisant le classificateur le plus proche moyen et la distance euclidienne.
2. Classer le point (9.6, 2) en utilisant le classificateur du plus proche voisin.
3. Classer le point (9.6, 2) en utilisant le classificateur K plus proche voisin lorsque K = 3.

Exercice 2 :

1. Calculer l'entropie d'un ensemble de deux événements également probables.
2. Calculer l'entropie d'un ensemble de quatre résultats possibles avec des probabilités $\{1/8, 3/4, 1/16, 1/16\}$
- Considérons les données suivants

a	b	c	d	Class
1	0	1	1	I
1	1	0	0	II
0	1	1	0	II
1	1	1	1	I

3. Construire l'arbre de décision à partir de la caractéristique a.
4. Construire un autre arbre à partir de la caractéristique sélectionnée en fonction de la mesure du contenu d'information.
5. Quelle est la différence entre les arbres? Pourquoi ?

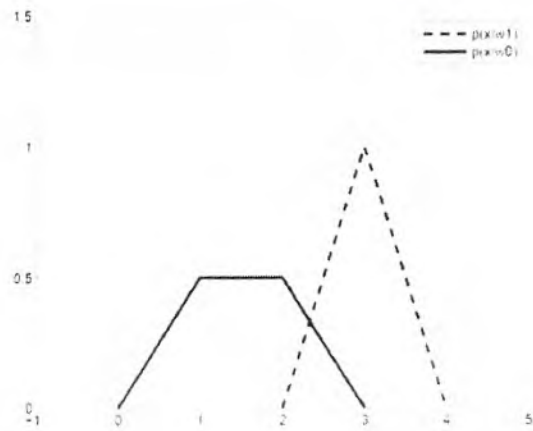
Exercice 3

Une base de données de 10 000 images et un utilisateur, qui ne s'intéresse qu'aux images contenant une voiture. On sait qu'il y a 500 de ces images dans la base de données. Un système de récupération automatique d'images récupère 300 images de voitures et 50 autres images de la base de données.

1. Déterminer la précision et le rappel du système de récupération.
2. Quelle est la relation entre la précision-rappel et les courbes ROC

Exercice 4

Considérons un système qui essaie de détecter la présence (ω_1) ou l'absence (ω_0) d'une classe. La figure ci-dessous illustre les distributions de probabilité conditionnelle de classe de l'observation x étant donné les deux cas.



1. Étant donné $P(\omega_0) = 3/5$ et $P(\omega_1) = 2/5$, trouvez les régions de décision R_0 et R_1 pour la règle de décision d'erreur minimale et calculez la probabilité d'erreur associée.