

Examen – Apprentissage**2h.****Polycopiés de cours non autorisés.**

Prenez soin de lire tous les exercices avant de commencer. La notation est donnée à titre indicatif.

Il y a 2 parties : **Partie 1** (14 pts) et **Partie 2** (6 pts). Chaque partie fera l'objet de copies séparées.**PARTIE 1****Exercice 1 : (6 pts)**

1. Rappeler les principes des méthodes d'apprentissage : Arbre de décision, Réseaux de neurones, k-PPV et Bayésien Naïf. (2 pts)
2. Quelles sont leurs ressemblances et leurs différences ? (2 pts)
3. En apprentissage supervisé, quelle est la différence entre un ensemble de test et un ensemble de validation ? (1 pt)
4. Citez trois méthodes d'évaluation des modèles d'apprentissage. (1 pt)

Exercice 2 : (8 pts)

@relation weather.symbolic

@attribute outlook {sunny, overcast, rainy}

@attribute temperature {hot, mild, cool}

@attribute humidity {high, normal}

@attribute windy {TRUE, FALSE}

@attribute play {yes, no}

@data

overcast, hot, high, FALSE, yes

rainy, mild, high, FALSE, yes

rainy, cool, normal, FALSE, yes

overcast, cool, normal, TRUE, yes

sunny, cool, normal, FALSE, yes

rainy, mild, normal, FALSE, yes

sunny, hot, high, FALSE, no

sunny, hot, high, TRUE, no

rainy, cool, normal, TRUE, no

sunny, mild, normal, TRUE, yes

sunny, mild, high, FALSE, no

overcast, mild, high, TRUE, yes

overcast, hot, normal, FALSE, yes

rainy, mild, high, TRUE, no

On dispose du fichier ci-dessus possédant une variable de classe PLAY matérialisant les conditions d'entraînement d'un joueur de tennis. On découpe l'ensemble en 2 : D1 et D2. **D1 contient les 9 premiers objets**, et D2 contient les **5 derniers**.

- 1- A quoi correspond le type de fichier ci-dessus ? Quel logiciel l'utilise ? (0,5 pt)
- 2- Construire le modèle bayésien naïf en utilisant D1 et en appliquant la formule de Laplace (voir ci-dessous) (2 pts)
- 3- L'ensemble D2 va être utilisé pour tester le modèle bayésien naïf. Déterminer la classe des 5 objets de D2. (2 pts)
- 4- Donner la matrice de confusion sur D2; (1 pt)
- 5- Calculer le taux d'erreur apparente de la méthode avec D2; (1 pt)
- 6- Calculer la précision de la classe 'yes' sur D2; (0,5 pt)
- 7- Calculer le rappel pour la classe 'no' sur D2; (0,5 pt)
- 8- On souhaite calculer la précision totale du modèle bayésien en tenant compte du poids de chacune des classes dans D2 (0,5 pt)

A_i : une valeur de l'attribut A

N_{ic} : Nombre d'objets ayant la valeur A_i dans la classe c

N_c : Nombre d'objets de la classe c

k : nombre de valeurs de l'attribut A

p : probabilité a priori

m : paramètre

$$\text{Original: } P(A_i | C) = \frac{N_{ic}}{N_c}$$

$$\text{Laplace: } P(A_i | C) = \frac{N_{ic} + 1}{N_c + k}$$

$$\text{m-estimate: } P(A_i | C) = \frac{N_{ic} + mp}{N_c + m}$$