

Travaux Dirigés
Intelligence artificielle, apprentissage automatique et big data
 Licence "Agriculture de Précision" – L3

Exercice 1 (Arbres de décision)

Soient des données contenues dans le tableau ci-dessous. Les attributs sont "Ensoleillement", "Espèce" et "Hydratation". La classe est "Maladie".

| Ensoleillement | Espèce | Hydratation | Maladie |
|----------------|--------|-------------|---------|
| oui | E1 | non | A |
| non | E1 | oui | A |
| oui | E2 | non | B |
| oui | E2 | oui | B |
| non | E2 | oui | C |
| oui | E3 | non | aucune |
| oui | E3 | oui | D |
| non | E3 | oui | E |
| non | E3 | non | aucune |

- 1) D'une façon naïve (sans utiliser la fonction Gain), proposer un arbre de décision pour l'attribut classe « Maladie »
- 2) Soit la formule de gain :

$$\text{Gain}(\mathcal{X}, a_j) = H(\mathcal{X}) - \sum_{v \in \text{valeurs}(a_j)} \frac{|\mathcal{X}_{a_j=v}|}{|\mathcal{X}|} H(\mathcal{X}_{a_j=v}) \quad H(\mathcal{X}) = - \sum_{i=1}^{i=n} p_i \log_2 p_i$$

Calculer le gain de chaque attribut du tableau donné.

Exercice 2 (KNN)

Soient des données sur des types de maladies relatifs à des plantes :

| Nom | Propriété 1 | Propriété 2 | Propriété 3 | Propriété 4 | Maladie |
|----------|-------------|-------------|-------------|-------------|----------|
| Plante 1 | 9 | 4 | 7 | 10 | Maladie1 |
| Plante 2 | 9 | 3 | 4 | 7 | Maladie2 |
| Plante 3 | 10 | 6 | 5 | 9 | Maladie2 |
| Plante 4 | 2 | 8 | 8 | 3 | Maladie1 |

Soit une plante dont les propriétés sont :

| Nom | Propriété 1 | Propriété 2 | Propriété 3 | Propriété 4 | Maladie |
|----------|-------------|-------------|-------------|-------------|---------|
| Plante A | 8 | 6 | 6 | 6 | ? |

- 1) Déterminer la maladie de la « Plante A » en exploitant KNN (les k plus proches voisins), en choisissant le paramètre K adapté.

Exercice 3 (Kmeans)

Soit l'ensemble D des entiers suivants : $D = \{2, 5, 8, 10, 11, 18, 20\}$

On veut répartir les données de D en trois (3) clusters, en utilisant l'algorithme Kmeans. La distance d entre deux nombres a et b est calculée ainsi : $d(a, b) = |a - b|$ (la valeur absolue de a moins b)

- 1) Appliquez Kmeans en choisissant comme centres initiaux des 3 clusters respectivement : 8, 10 et 11. Montrez toutes les étapes de calcul.