Méthodes et Outils de Traitements d'images Avancés
TELECOM Nancy 2ème Année

TD1: K-means – Decision Tree

Exercice n°1 : K-means avec lot de données Iris

1.1. Appliquer l'algorithme des K-means (***kmeans*** - paramètres par défauts) sur le lot de données iris en utilisant les attributs *longueur et largeur des pétales*. Visualiser les clusters obtenus.

1.2. Calculer la matrice de confusion issue de cette partition (***confusionmat***). Quel problème peut-on rencontrer ?

1.3. Evaluer les performances du clustering obtenu :

1.1.3.1. Calculer l'entropie, la pureté.

1.1.3.2. Calculer la justesse (accuracy : taux de classification / classes / total)

1.1.3.3. Calculer la précision, le rappel et la F-Value.

1.4. Modifier les conditions initiales pour fixer les seuils initiaux à :

C= [0 0; 0 0; 0 0] / C= [0 0; 2.5 5; 5 10] / C= [0 0; 2.5 1.5; 5 2.5]

Que conclure ?

1.5. Effectuer le clustering en utilisant la distance de Manhattan (option ***'Distance','cityblock'***).

1.6. Effectuer la classification en utilisant 3 puis 4 critères caractérisant les iris. Calculer les indicateurs de performance (notamment l'entropie) que conclure ?

Exercice n°2 : DT avec lot de données Iris

2.1. Utiliser le module « Classification Learner ». Créer une nouvelle session en important les données UCI « iris.txt ». Configurer le module pour travailler sans validation.

2.2. Entrainer un arbre de décision simple (tree) sur le lot de donnée iris avec les paramétrages par défaut avec toutes les variables d'entrée.

2.3. Visualiser la matrice de confusion correspondante et donner les mesures d'évaluation (justesse, rappel, précision, …)

2.4. Exporter l'arbre de décision crée et le visualiser avec la commande view(tree.classificationTree, 'Mode','graph').

2.5. Modifier le paramétrage de la session pour effectuer une validation croisée 2/3 – 1/3 puis 9/10 – 1/10). Evaluer les modèles obtenus, que conclure ?

2.6. Elaguer les arbres obtenus à différent niveau. Visualiser les arbres obtenus et tester leur efficacité (Option avancées).

2.7. Modifier le critère de sélection de caractéristique avec l'option avancée ***'SplitCriterion', 'Maximum Deviance'*** pour travailler avec l'entropy (Gini par défaut). Evaluer la performance du modèle obtenu.

2.8. Effectuer la classification en utilisant 3 puis 2 critères caractérisant les iris. Calculer les indicateurs de performance, que conclure ?