

TD2: Decision Tree - Etude du lot de données « Radiata Pin Singularities »

1. Charger le lot de données « Radiata Pin » en utilisant le « **classification learner** »

Evaluer la classification obtenue avec le paramétrage par défaut pour des arbres « complex » « medium » et « simple ».

Visualiser l'arbre obtenu avec la commande : **view(tree.ClassificationTree, 'Mode','graph')** après avoir exporté le modèle compact. Quels sont les attributs réellement utilisés (variables d'entrée).
2. Analyser la matrice de confusion correspondante et évaluer la classification obtenue en vous appuyant sur la justesse, la précision et le rappel.
3. Elaguer l'arbre obtenu à différent niveau *i* : (**options avancées**). Visualiser les arbres obtenus et tester leur efficacité.
4. Exporter le modèle et le tester l'arbre obtenu sur le lot de données de test : '**lot_TST_Arbo_Radiata.xls**' à l'aide de la commande (Yp=tree.predictFcn(X). Evaluer la classification à l'aide de la matrice de confusion, que conclure ?
5. Créer une nouvelle session pour effectuer une validation croisée 2/3 – 1/3. Evaluer la classification obtenue. Que conclure ?
6. L'étude des variables d'entrées montrent que seuls les 11 attributs suivants sont pertinents : DX/DY, LN_RE, GD_AXE, PT_AXE, LR_RE, C1+C3, C1, C3, C4, SURF, ORIENT. Refaire la classification avec ces attributs en entrée. Evaluer la classification. Que conclure ?
7. Elaguer l'arbre obtenu à différent niveau *i* (**options avancées**). Visualiser les arbres obtenus et tester leur efficacité.
8. Tester l'arbre obtenu sur le lot de données de test : '**lot_TST_Arbo_Radiata.xls**'. Que conclure ?
9. Effectuer une validation croisée 2/3 – 1/3. Evaluer la classification obtenue.
10. Mêmes questions que 5 à 8 mais en utilisant les attributs sélectionnés par des méthodes automatiques (SFFS¹/SBFS¹/FRIFS²): PT_AXE, DX/DY, LR_RE, C1, C4, C1+C3.
11. Les experts proposent une Structure Arborescente Hiérarchique³ décrite dans le document Arbo_Radiata.pdf. Créer un classificateur de type DT pour chaque niveau. Tester sa validité. Que conclure ?

¹ Pudil, P., Novovicova, J., and Kittler, J. (1994). Floating search methods in feature selection, *Pattern recognition letters* 15:1119-1125.

² Schmitt, E., Bombardier, V., and Wendling, L. (2008). Improving Fuzzy Rule Classifier by Extracting Suitable Features From Capacities With Respect to the Choquet Integral, *IEEE trans. On Systems, man and Cybernetics-Part B: Cybernetics*, 38 (5) :1195-1206

³ Bombardier V., Mazaud C., Lhoste P. Vogrig R. (2007) Contribution of Fuzzy Reasoning Method to knowledge Integration in a wood defect Recognition System. *Computers in Industry Journal* 58:355–366.