

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

1. Charger le lot de données « Lot_App_Arbo_Radiata.xls » en utilisant le « **classification learner** ». La colonne A correspond aux 9 classes de sorties, celles de G à Z contiennent les 20 attributs d'entrée.

Créer une **session sans validation** et générer les arbres « T_coarse » et « T_fine ». Analyser la matrice de confusion correspondante et évaluer le modèle de classification obtenu en vous appuyant sur la justesse, la précision et le rappel.

Visualiser les arbres obtenus avec la commande : **view(T_x.ClassificationTree, 'Mode', 'graph')** après avoir exporté le modèle compact. Quels sont les attributs réellement utilisés (variables d'entrée).
2. Générer un arbre en contrôlant le niveau de profondeur i : (**options avancées**) et tester son efficacité. Choisir un niveau de profondeur en justifiant votre choix. Le niveau choisi sera conservé pour la suite.
3. Exporter le modèle obtenu et le tester sur le lot de données de test '**lot_TST_Arbo_Radiata.xls**' (à importer avec « **Import Data** » : Y en format Cells et X en format Table) à l'aide de la commande **Yp=tree.predictFcn(X)**. Evaluer la classification à l'aide de la matrice de confusion (**confusionmat(Y, Yp)**). Que conclure ?
4. Créer une nouvelle session pour effectuer une **validation croisée 4/5 – 1/5**. Evaluer l'efficacité du modèle de classification obtenu. Comparer à ceux obtenus dans les questions précédentes. Que conclure ?
5. 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 une session **sans validation** avec ces attributs en entrée. Evaluer la classification. Que conclure ?
6. Exporter et tester l'arbre obtenu sur le lot de données de test : '**lot_TST_Arbo_Radiata.xls**'. Que conclure ?
7. Créer une nouvelle session avec une **validation croisée 4/5 – 1/5**. Evaluer la classification obtenue.
8. Mêmes questions que 5 à 7 mais en utilisant les attributs sélectionnés par des méthodes automatiques (SFFS¹/SBFS¹/FRIFS²): PT_AXE, DX/DY, LR_RE, C1, C4, C1+C3.
9. Les experts proposent une Structure Arborescente Hiérarchique³ décrite dans le document Arbo_Radiata.pdf. Créer une session avec **validation croisée** pour chaque nœud de décision. Tester l'efficacité individuellement et estimer la classification globale. 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.