

Apprentissage Inductif Appliqué

Travaux pratiques : Utilisation de PEPITo

Dans le cadre du cours d'apprentissage inductif appliqué, il vous est demandé de répondre aux questions suivantes en vous aidant du logiciel PEPITo et de la base de données Omib

1. Construisez un arbre de décision en utilisant un ensemble d'apprentissage, de test et d'élagage tous de taille 1000 (sans recouvrement) afin de modéliser l'attribut *SECURITY* à l'aide des 6 attributs Pu, Qu, Vl, Vinf, Pl, Xinf.
 - (a) Comparez avec la méthode du plus proche voisin et celle des réseaux de neurones (en faisant varier les paramètres de ces méthodes). Comment les trois méthodes se comportent-elles (précision, complexité des modèles) si vous ajoutez un attribut non pertinent (p.ex. object-id) ou si vous supprimez un attribut pertinent (p.ex. Pu ou Xinf).
 - (b) Comment varie l'erreur si vous faites varier l'ensemble d'apprentissage (d'abord en gardant la même taille, puis en prenant une taille supérieure ou inférieure, en gardant les mêmes ensemble de test et d'élagage et en s'assurant qu'il n'y ait pas de recouvrement) ?
 - (c) Idem en faisant varier l'ensemble de test et/ou l'ensemble d'élagage.
2. Comparez les méthodes de pré et post-élagage. Le pré-élagage vous permet-t-il de diminuer l'erreur ? (faites plusieurs tests avec des ensembles d'apprentissage différents et des valeurs de *alpha* différentes)
3. Comparez la méthode de régression linéaire et les réseaux de neurones en termes de temps de calcul et de précision pour la prédiction de la la variable de sortie *CCT-SBS* (faites plusieurs tests pour chaque méthode). Quelle est la plus efficace ? Quid si vous ajoutez un attribut non pertinent, ou si vous enlevez un attribut pertinent. Pouvez-vous observer le sur-apprentissage avec les réseaux de neurones ?
4. Quelle méthode de classification (vue au cours) vous semble la plus efficace pour prédire la valeur de l'attribut *SECURITY* ?
5. Mis à part les attributs *SECURITY* et *CCT-SBS*, quel(s) attribut(s) est(sont) le(s) plus important(s) pour déterminer la valeur de sortie *SECURITY* ? Justifiez votre réponse.

Il vous est également demandé de rendre un rapport qui décrit de manière précise les essais que vous avez réalisés, les résultats obtenus les principales conclusions que vous en aurez tiré sur le travail que vous aurez effectué. Vous pouvez le rendre sous format papier (boîte postale 15) ou sous format électronique (formats acceptés : doc, pdf, txt, jpeg, png) par mail à S.Hiard@ulg.ac.be.

DATE LIMITE DE REMISE DES TRAVAUX : 17 novembre 2008

Bon travail...