Projet Intelligence Artificielle – Printemps 2020

Task 1 :

Il s’agissait de créer un arbre de décision basé sur l’algorithme ID3 donné dans les solutions des exercices. Il fallait tout d’abord convertir le fichier CSV en tableau de données avec notre fonction csv\_reader dans task\_1/csv\_reader.py. Les entrées du tableau sont des string, car cela permettait un affichage plus simple par la suite. Puis nous avons créer l’arbre en s’aidant des données d’entrainement.

*Statistiques Tabin*

Task 2 :

Pour pouvoir prédire la précision de prédictions correctes, nous avons utilisé les données de test, et comparions le résultat target prédit par l’algorithme id3 (1 ou 0) par rapport à la « vraie » valeur présente dans les données. Nous obtenions 56.25%

Task 3 :

Task 4 :

Task 5 :

Nous devions modifier l’algorithme ID3 pour qu’il puisse traiter des données continues, c’est-à-dire créer un arbre avec des nœuds binomiaux : à gauche les données plus petites qu’une certaine valeur de séparation, à droite les valeurs plus grandes ou égales à cette valeur.

Tout d’abord il fallait séléctionner la combinaison « attribut = valeur » qui réduisait au maximum l’entropie, et cette valeur devenait la valeur de séparation (fonction construit\_arbre\_recur dans task\_5/id3\_continuous).

Puis il fallait séparer les données en 2 nouveaux set de données qui correspondent à chaque nouveau nœud (fonction partitionne dans task\_5/id3\_continuous). Il fallait aussi ajuster le tableau attributs en fonction des valeurs qui seraient conservées dans le nouveau nœud. Puis enfin nous créons les 2 enfants, et finalement nous créons l’entité NœudDeDecision\_continuous.

Dans NœudDeDecision\_continuous, il fallait modifier l’affichage : dans classifie, il fallait séparer les données dans l’enfant de gauche ou de droite, puis modifier l’affichage (< et >=).

Une fois cette arbre avancé fini, nous pouvions calculer la précision des prédictions. Nous obtenons 51.249999999. (Légèrement en dessous de la précision trouvée au task 2, nous nous attendions à trouver une prédiction plus élevée).