

0.1 Arbre de décision

0.1.1 Définition

L'arbre de décision est une technique d'apprentissage automatique très utilisée, l'idée consiste à construire un modèle(sous forme d'arbre) à partir de données d'entraînement, ensuite, l'utiliser afin de classer de manière automatique, de nouvelles instances.

Afin de construire un arbre de décision, il existe une variété d'algorithmes, avec la même idée de base mais un traitement différent, nous pouvons donc avoir, avec le même ensemble de données d'apprentissage, des arbres de décision différents créés par chacun de ces algorithmes. A partir de là, nous devons donc définir quelle est la mesure qui permet de comparer la qualité des arbres de décision ?

Plusieurs mesure existes, nous pouvons citer :

- Précision prédictive
cela fait référence à capacité de l'arbre à prédire *correctement* la classe de nouvelles instances (autres que les données d'apprentissage).
Cette mesure est calculée à partir d'un ensemble de données teste, en utilisant la formule suivante :
$$\text{Précision de l'arbre (\%)} = \frac{\text{Nombre d'instance bien classé}}{\text{Nombre total d'instance}}$$
- Vitesse
Cette mesure fait référence aux coûts (en terme de temps) de calcul impliqués en générant et en utilisant l'arbre.
- Robustesse
Capacité de l'arbre à s'adapter à des situations "moins bonne", tel que :
 - Gestion d'une instance à valeur manquante
 - Gestion d'attributs mixtes (continus et discrets)
- Simplicité et interprétabilité
La simplicité fait référence à la taille de l'arbre construit, plus cette taille est réduite, plus sa simplicité et son interprétabilité augmente,
- Evolutivité
Sa capacité à maintenir de bonne performance lorsque le volume de données augmente

0.1.2 Quel est le meilleur critère ?

L'idéal pour un arbre de décision est d'assurer à 100% tous les critères citer en haut, chose qui, en réalité est très difficile(voire impossible), car certain critère ne sont obtenu qu'au détriment d'autres, un arbre de décision parfaitement simple et réduit aura certainement une faible précision, contrairement à un arbre très complexe et grand, qui, malgré sa parfaite précision, sera mauvais en terme de temps.

Ces critères sont choisis selon le contexte et la nature du traitement voulu à travers l'arbre de décision.

Prenons l'exemple d'une situation où l'analyse doit se faire de manière régulière, nous pouvons considérer l'arbre optimal comme étant le plus rapide (Tout en précisant un certain seuil de précision bien entendu), contrairement à une analyse qui se fait de manière occasionnelle, dont le résultat ne peut être erroné, là l'arbre optimal serait le plus précis, en dépit de sa taille et de son temps d'analyse.