

Rapport

****//Préparations des données****

*

1. Ces données comportent 14 attributs
2. 4 classes différentes catégorises les données
3. **__Utiliser read file puis choisir chaque classe que l'on veut étudiée__**

Classe 0 : 674

Classe 1 : 908

Classe 2 : 472

Classe 3 : 244

4. Réponses dans le code 'code_questions.py' = Inséparables linéairement

5. Pour l'arbre de décision le one-hot peut-être utile mais pas forcément
Pour le réseaux de neurones il est on pense plus intéressant que pour l'arbre
La normalisation quant à elle sera utile dans les réseaux de neurones

6. L'intêret est de pouvoir comparer nos valeurs de test et les vérités du terrain afin d'entraîner correctement notre modèle donc en évitant le sur-apprentissage grâce a l'early-stopping qui permet d'éviter cela justement en séparant nos données en un jeu de test et un jeu d'entraînement.

****//2 Mise en oeuvre des modèles****

****//2.1 Arbre de Décision****

1. Une des remarques que l'on peut faire et que les quartiles correspondent a une division égal de la population total de chaque attribut de notre table de données . Ce ne permet donc plus de se focaliser sur une colonne entière mais sur une partie de celle ci et donc de pouvoir subdiviser encore notre colonne pour avoir un grain plus fin et donc rechercher le meilleur gain sur un endroit précis de notre table de données.

****//2.2 Réseaux de neurones artificiels****

****//3 Analyse des modèles****

A retrouver dans analysis.py

DT4:

Modèle M

Classes C1 C2 C3 C4

Accuracy [] [] [] []

Precision [] [] [] []

Recall [] [] [] []

F1-Score [] [] [] []

DT5:

Modèle M

Classes C1 C2 C3 C4

Accuracy [] [] [] []
Precision [] [] [] []
Recall [] [] [] []
F1-Score [] [] [] []

DT6:

Modèle M

Classes C1 C2 C3 C4

Accuracy [] [] [] []

Precision [] [] [] []

Recall [] [] [] []

F1-Score [] [] [] []

****//4 Le meilleur modèle****

1. On a tout d'abord chercher le meilleur modèle pour chaque catégorie (DT,relu,tanh)
Dans ce cas là le meilleur pour DT est DT5, pour Relu c'est 10-8-6 (dur bataille avec 10-8-4)

Et dans les tanh c'est sans débat le 10-8-6

Maintenant si l'on doit comparer les 3

Dans toutes les catégories F1/Recall/Precision/Accuracy

Tanh est bien devant le relu et le DT5 est pas mal derrière dans toutes ces catégories

Pour la comparaison des différent modèles on a juste fait la somme des valeurs dans chaque catégorie comparer a un autre modèle dans les même catégorie et si la somme de l'un est plus grande que l'autre alors c'est lui le meilleur modèle des 2

2. On peut partir du principe que les réseaux de neurones sont plus précis et qu'ils ont étaient créer pour apporter de la précision dans beaucoup de domaine donc on dira que les réseaux de neurones seront plus adapté.