

TP 2 – Clustering et classification

le 3 novembre 2020

- terminez le TP de la séance précédente (si c'est le cas) et retrouvez le code R écrit à cette occasion.

- En utilisant R¹ appliquez les fonctions de clustering (classification non-supervisée) y compris DBSCAN sur les données `mammal.dentition` qui existent dans la base R dans la librairie `cluster.datasets`². Il faut, naturellement, ignorer la première colonne. Indiquez quelle méthode (algorithme / distance) produit les meilleurs résultats.

- Classifiez manuellement les animaux, éliminez (s'il y a) les classes avec une seule observation, puis utilisez la fonction de classification par la méthode k-NN. Dressez la matrice de confusion.

- Quelle est le meilleur choix pour la valeur k ?

Compte-rendu sous la forme d'un unique fichier (code R commenté ou Python ou archive .zip ou .rar avec code .r et compte-rendu en .pdf) à rendre sur Campus. Ce qui intéresse c'est l'interprétation des résultats obtenus.

1. Vous pouvez également utiliser Python
2. un fichier .csv et disponible pour ceux qui utilisent Python.