Clément Baudry 24/05/2022

AWOUKOU Maxime

Rapport KNN

# **L’algorithme du KNN**

Nous avons récupéré l’algorithme du KKN, du TD portant sur ce sujet, et nous l’avons adapté au problème posé.

Une première méthode DistanceMatrix renvoie une matrice de distances en fonction des vecteurs contenus dans le fichier qu'on utilise pour entrainer l’algorithme.

Une seconde méthode LoadDatas permet de mettre sous forme, de liste l’ensemble des vecteurs d’un fichier donné. Elle permet aussi de convertir les valeurs en float avec la méthode ConvertFloat et de distinguer les fichiers de test et les fichiers pour les tests finaux (sans valeurs 0 et 1 à la fin).

Enfin, la méthode FindClosestValue cherche les k voisins les plus proches et retourne la valeur choisie en conséquence. Cette

# **Recherche du meilleur K**

La méthode ConfusionMatrix permet d’observer les vrais positifs et les faux positifs ainsi que les faux positifs et les faux négatifs en fonction des valeurs de références et des valeurs trouvées. De plus, la méthode Eval nous permet d’observer le pourcentage de réussite total.

Pour la recherche du meilleur K, nous avons utilisé une partie du fichier data.txt pour entrainer l’algorithme et une autre partie du fichier pour tester notre algorithme. Nous avons fait ces tests, de nombreuses fois avec des K diffèrent et nous obtenions en meilleur K : (23,29,87) avec un pourcentage de réussite à environ 84%.

Puis nous avons réalisé le test avec le fichier preTest.txt et cette fois c’est la valeur 29 qui est la plus ressortie.

Nous avons donc choisi un k égal à 29 pour le test final.