



Algorithme des k plus proches voisins

Présentation : L'algorithme des k plus proches voisins est un algorithme d'apprentissage automatique (*machine learning*).

Ce type d'algorithmes, qui a vu le jour aux USA à la fin des années 50, connaît depuis le début des années 2000 un regain d'intérêt car il permet une bonne exploitation des données présentes sur internet.

Ce type d'algorithme permet à partir d'un certain nombre de données classées dans un ensemble E d'en intégrer une autre.

Principe

On possède un ensemble E de données qui sont classées. On connaît leur classe et leurs caractéristiques.

On s'intéresse à une donnée x dont on connaît les caractéristiques mais pas la classe.

On va calculer la distance entre la donnée x et chacune des données de l'ensemble E.

On retient les k données de l'ensemble E qui sont les plus proches de x.

On attribue à x la classe des données les plus fréquentes parmi les données retenues.

Exemple

Un exemple un peu morbide est lié au Titanic. On classe les passagers du Titanic en deux catégories : Morts ou survivants.

Dans notre ensemble de données connu, un élément appartient donc à mort ou survivant et possède différentes caractéristiques : Age, sexe, numéro de cabine, pont de la cabine, milieu social...

L'algorithme des k plus proches voisins va permettre en prenant un individu dont on ne connaît pas la classe de le classer dans la catégorie mort ou survivant.

Pour calculer ces distances, on rappelle la formule mathématique de la distance entre deux points :

La distance entre deux points dont les positions sont $(x_A; y_A)$ et $(x_B; y_B)$ est donnée par la formule :

$$dist(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}$$

Si on a trois caractéristiques numériques pour chaque point : $(x_A; y_A; z_A)$ et $(x_B; y_B; z_B)$ on obtient :

$$dist(A, B) = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$$

Cela se généralise à quatre descripteurs, cinq descripteurs ... et N descripteurs.

Un algorithme de visualisation du principe.

[Visualiser le principe de l'algorithme](#)