

Algorithme des k plus proches voisins

1) Apprentissage automatique (Machine learning)

L'algorithme des k plus proches voisins (**K-NN** en anglais pour K-nearest neighbors) appartient à la famille des algorithmes d'apprentissage automatique (machine learning).

Machine Learning : Donner la possibilité à l'ordinateur d'apprendre sans être programmé.

L'apprentissage automatique est un domaine qui se développe beaucoup. De nombreuses entreprises et organisations utilisent les données concernant leurs « clients » afin de « nourrir » des algorithmes de machine learning qui permettront d'en savoir toujours plus sur eux et ainsi de mieux cerner et prévoir les « besoins » et les « comportements ».

La technologie de reconnaissance faciale par exemple permet de marquer des photos. La technologie de reconnaissance optique des caractères (OCR) convertit les images du texte en caractères mobiles. Les voitures autonomes utiliseront l'apprentissage automatique pour naviguer.

En tant que domaine informatique, l'apprentissage automatique est étroitement lié aux statistiques mathématiques.

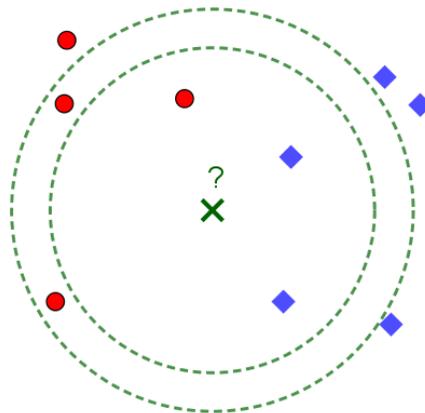
2) K-NN – k-plus proche voisin

L'algorithme de k-plus proche voisin est un modèle de reconnaissance de modèle qui peut être utilisé pour la **classification**. Le but de cet algorithme est d'assigner un nouvel objet à la classe la plus commune parmi ses k plus proches voisins.

Regardons un exemple : dans le diagramme ci-dessous, il y a des objets « **carré bleu** » et des objets « **disque rouge** ». Lorsqu'un nouvel objet est ajouté à l'espace - dans ce cas, la croix verte - nous voulons que l'algorithme d'apprentissage classe cette croix dans une certaine classe.

Pour $k = 3$, l'algorithme trouvera les trois plus proches voisins de la croix afin de la classer soit dans la classe des disques, soit dans la classe des carrés. Dans notre diagramme, les trois plus proches voisins de la croix sont un disque et deux carrés. Par conséquent, l'algorithme classera la croix les carrés.

Pour $k = 5$, l'algorithme classera la croix avec les disques.

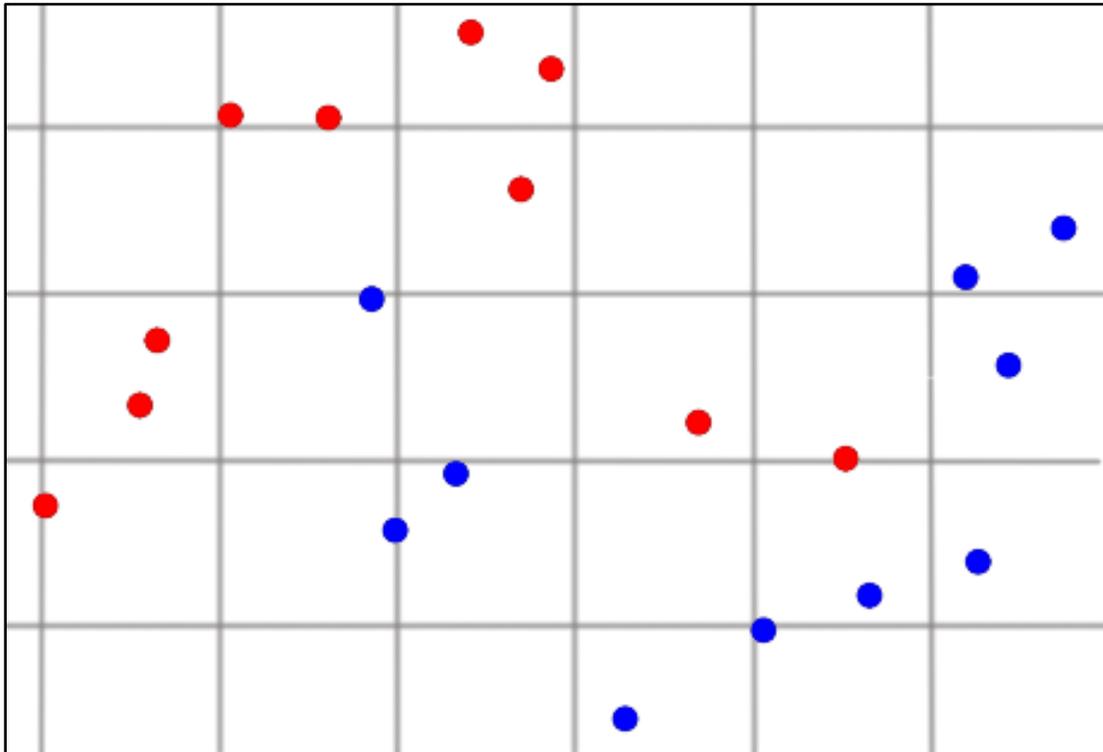


En utilisant cette méthode, on peut facilement en déduire les zones rouges et bleues, où les points qui se situeront dans la zone seront classés comme rouges ou bleus.

Exercice : tracer les zones rouges et bleues avec le modèle 5-NN

Les points qui se situeront dans la zone rouge seront classés comme rouges

Les points qui se situeront dans la zone bleue seront classés comme bleus



3) Principe de programmation d'un algorithme k-plus proche voisin

L'algorithme est construit en 3 étapes :

Etape 1 : Calculer les distances (par exemples les distances euclidiennes dans l'exemple ci-dessus)

Etape 2 : Trier les données calculées en fonction des distances.

Etape 3 : Déterminer les k plus proches voisins et classer en fonction de la fréquence du caractère.