kNN 鄰近分類演算法

將未知資料與已知類別的資料作比較

將每筆資料中的特徵值模擬成空間向 量

並計算未知資料與已知資料的距離，必選擇k個最近的已知實例

並以少數服從多數的法則，將未知資料歸類於k個鄰近樣本中最多的類別

優點：

可用於非線性問題

準確度高，對例外狀況不敏感

適用於多分類問題

缺點：

計算量大，每個點皆須計算距離

樣本不平衡時，容易造成分類偏差

k-mean 聚類演算法

將資料模擬成空間向量，隨機取得k個值（也就是類別），作為群聚中心，並計算每點到聚類中心之距離，並取平均得到新的群聚中心，最後重複上述動作，到所只是條件 或是群聚中心不在跳動

優點

聚類效果佳

計算時間短

易理解

缺點

對異常值非常敏感

初始中心選擇敏感，易收斂城區域解

需要預先設置k個分類，不易選定k值

對於分類：各類樣本不平橫，分類結果有重疊，非類結果無法以一條直線做分類，容易造成分類偏差，如甜甜圈形狀

收斂選擇

群體形狀不一與群體大小不同時 一造成分類偏差

K-Means是无监督学习的聚类算法，没有样本输出；而KNN是监督学习的分类算法，有对应的类别输出。KNN基本不需要训练，对测试集里面的点，只需要找到在训练集中最近的k个点，用这最近的k个点的类别来决定测试点的类别。而K-Means则有明显的训练过程，找到k个类别的最佳质心，从而决定样本的簇类别。