

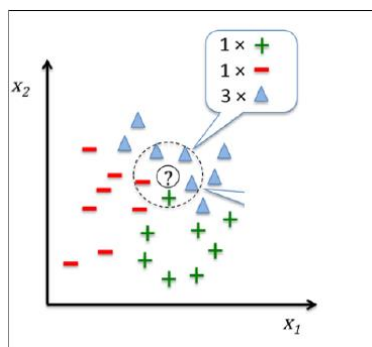
## KNN (K-近鄰演算法)

### 定義

監督式學習演算法，其演算法是根據距離相近的鄰居類別來判定自己的所屬類別。

步驟為：

- 一、確定  $k$  大小和距離度量 ( $k = 5$ )。
2. 對於測試集中的一個樣本，找到訓練集中和它最近的  $k$  個樣本。
3. 將這  $k$  個樣本的投票結果作為測試樣本的類別（多數決）。



- 通常  $K$  為奇數
- 除了多數決的方法外，還有以距離為基礎之權重方式。通常權重會是距離的倒數
- 當  $k=1$  的時候容易 Overfitting training data；而  $k$  很大的時候容易 underfitting training data
- 優點: 1. 對異常值不敏感 2. 參數簡單 3. 概念簡單易實現
- 缺點: 1. 距離的計算量龐大 2. 記憶空間需求量大

## 二、 參考資料

1. Class Handout, Lee, Chia-Jung professor, MDM64001, School of Big Data Management, Soochow University
2. K 近鄰 k-Nearest Neighbor (KNN) 演算法的理解  
<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/635163/>
3. [Machine Learning] kNN 分類演算法  
<https://medium.com/@NorthBei/machine-learning-knn%E5%88%86%E9%A1%9E%E6%BC%94%E7%AE%97%E6%B3%95-b3e9b5aea8df>