Machine Learning 筆記

1. 支援向量機(Support Vector Machine, SVM)

找出一個超平面(hyperplane)，使之將兩個不同的集合分開。以二維平面來說，我們希望能找出一條線能夠將兩種不同的點分開，而且我們還希望這條線距離這兩個集合的邊界越大越好。

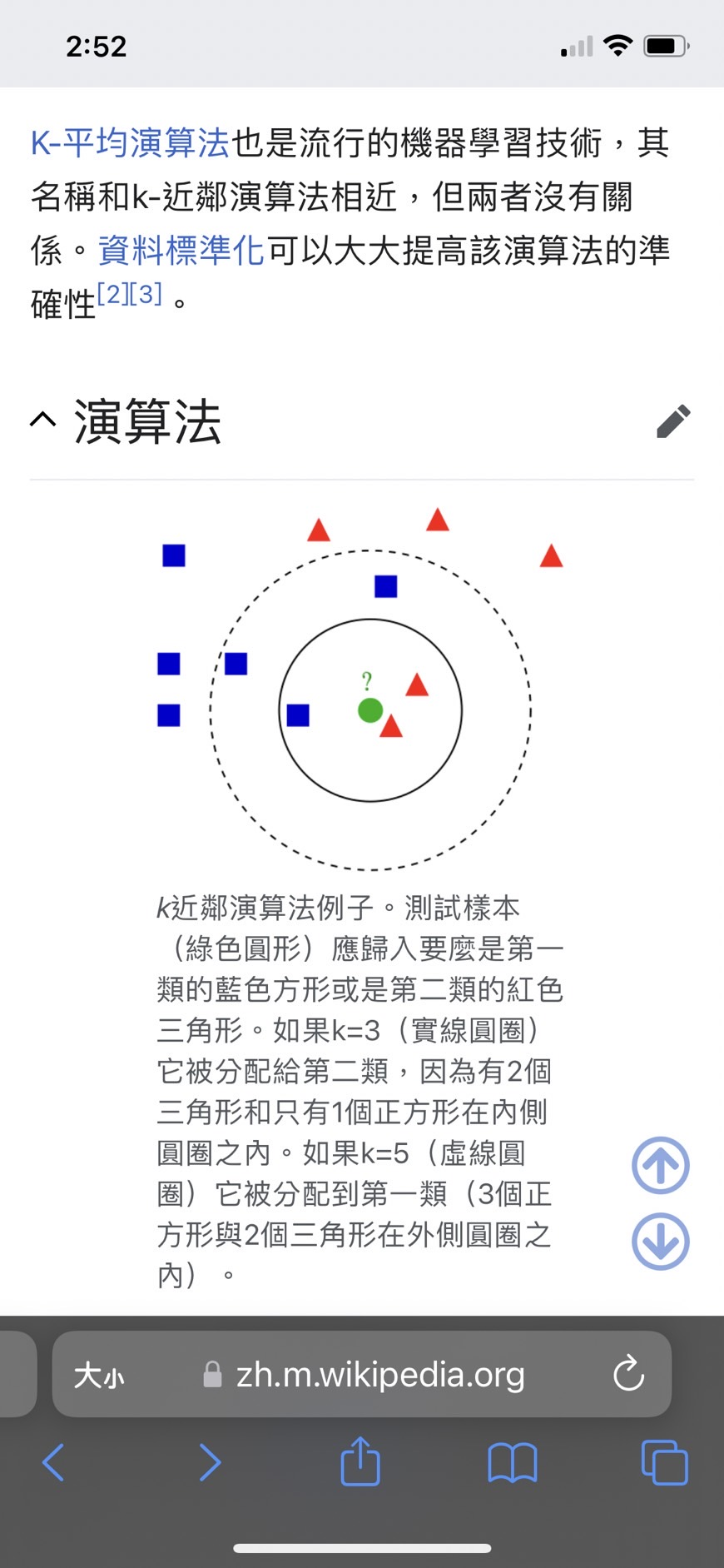
<https://mropengate.blogspot.com/2015/03/support-vector-machines-svm.html?m=1>

四種不同SVC分類器:

1. LinearSVC (線性)
2. kernel='linear' (線性)
3. kernel='poly' (非線性)
4. kernel='rbf' (非線性)

<https://ithelp.ithome.com.tw/m/articles/10270447>

1. K-近鄰演算法(K-nearest neighbors, KNN)



n\_neighbors = 5，代表某資料在定義標籤時會參考最靠近的五筆資料

n\_cluster 是用來設定「產生幾組」群，

例如：若 n\_cluster = 3，代表產生出來的標籤為0、1與2

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

<https://ithelp.ithome.com.tw/m/articles/10269826>

<https://youtu.be/UeyZX31VZE8>

1. 決策樹(Decision Tree, DT)

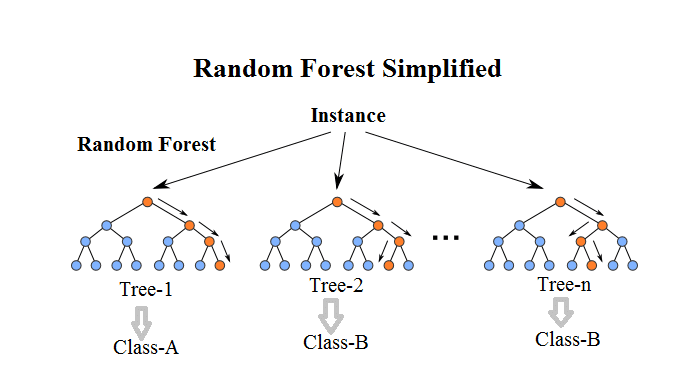
<https://medium.com/chung-yi/ml%E5%85%A5%E9%96%80-%E5%8D%81%E5%85%AD-%E6%B1%BA%E7%AD%96%E6%A8%B9-decision-tree-59e5fb6a0f56>



<https://www.kaggle.com/code/ankur02/mobile-price-classifier-eda>

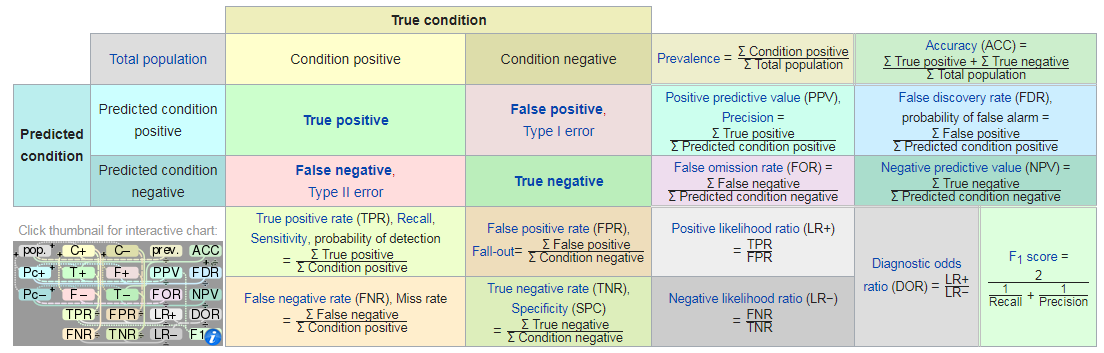
1. 隨機森林(Random Forest, RF)

Random Forest = Bagging + Decision Tree



<https://medium.com/chung-yi/ml%E5%85%A5%E9%96%80-%E5%8D%81%E4%B8%83-%E9%9A%A8%E6%A9%9F%E6%A3%AE%E6%9E%97-random-forest-6afc24871857>

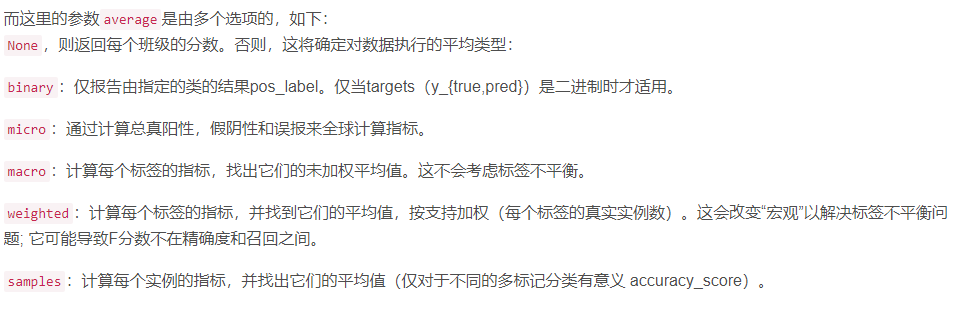
混淆矩陣(Confusion matrix)



<https://www.ycc.idv.tw/confusion-matrix.html>

**Precision看的是在預測正向的情形下，實際的「精準度」是多少**

**Recall則是看在實際情形為正向的狀況下，預測「能召回多少」實際正向的答案**



<https://blog.csdn.net/datongmu_yile/article/details/81750737>

<https://www.kaggle.com/code/arpit2807/multi-class-classification-for-beginner>s

其他

正規化(已全是數值)

標準化

<https://ithelp.ithome.com.tw/articles/10216967>

<https://www.youtube.com/watch?v=3GxT8n0ShsU&t=184s>

相關係數刪欄位

df.corr(method ='pearson')

**參數：**  
**method:**  
pearson ：標準相關係數  
kendall ：Kendall Tau相關係數  
spearman ：矛兵等級相關

<https://www.kaggle.com/code/ankur02/mobile-price-classifier-eda>