



機器學習模型：隨機森林 Random Forest Algorithm (RF)

如何利用RF來預測股票的漲跌？



RF 的優點？

- ❑ 不太需要data preprocessing (處理資料)
- ❑ 沒有太多Hyper-parameters (預測參數)
- ❑ 看的懂RF思考的思路
- ❑ 不容易Overfitting (過擬合)



RF 究竟是Classifier (分類) 還是Regressor (回歸) ?

- 都可以喔！全名叫做
 - Random Forest Classifier (分類)
 - Random Forest Regressor (回歸)



Decision Tree (決策樹)

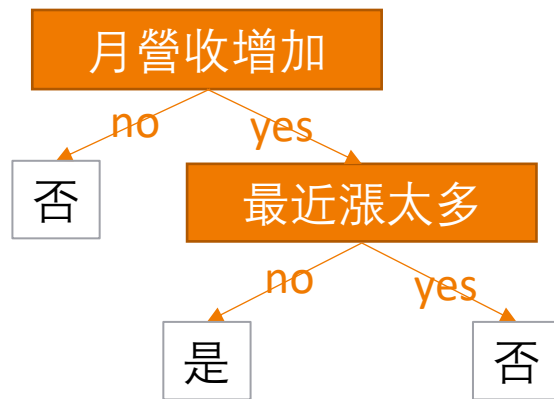


投資久了
會形成固定的策略



Decision Tree (決策樹)
用電腦來生成此策略

是否要投資該股票？





如何建構Decision Tree (決策樹)?

- ❑ ID3 - Iterative Dichotomiser 3 (疊代二元樹3代)
 - ❑ Entropy 熵
 - ❑ Information Gain 熵的變化量



熵 Entropy (E)

□ 一個數字，代表該系統混亂的程度





Information Gain (G)

□ 熵 entropy 降低的量





Splitting (決策樹分割)

□ 今天出門要不要帶傘？

features	正確率 (歷史數據)	熵
就是帶傘	60%	0.97
擲銅板	49%	0.99
擲筊	51%	0.98
看窗外	75%	0.81
看天氣預報	90%	0.46

熵最低！

要

365天的數據

看天氣預報

甬

要

200天的數據

165天的數據



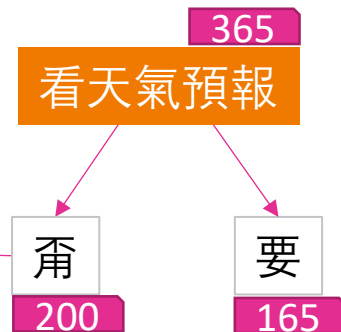
Splitting (決策樹分割)

200天的數據

features	正確率	熵
就是帶傘	90%	0.46
擲銅板	51%	0.99
擲筊	49%	0.98
看窗外	99%	0.08

氣象預報說晴天、窗外下雨

是否帶傘？





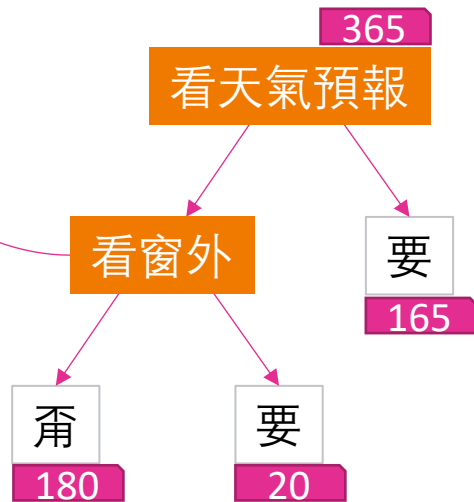
Splitting (決策樹分割)

200天的數據

features	正確率	熵
就是帶傘	90%	0.46
擲銅板	51%	0.99
擲筊	49%	0.98
看窗外	99%	0.08

氣象預報說晴天、窗外下雨

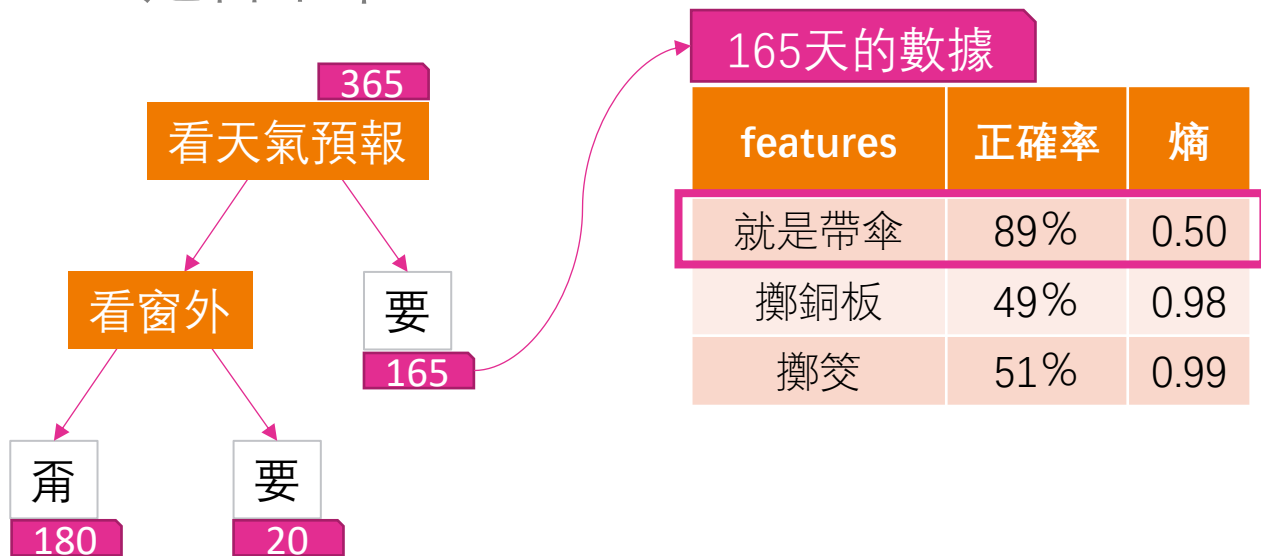
是否帶傘？





Splitting (決策樹分割)

是否帶傘？





Decision Tree (決策樹) 的缺點

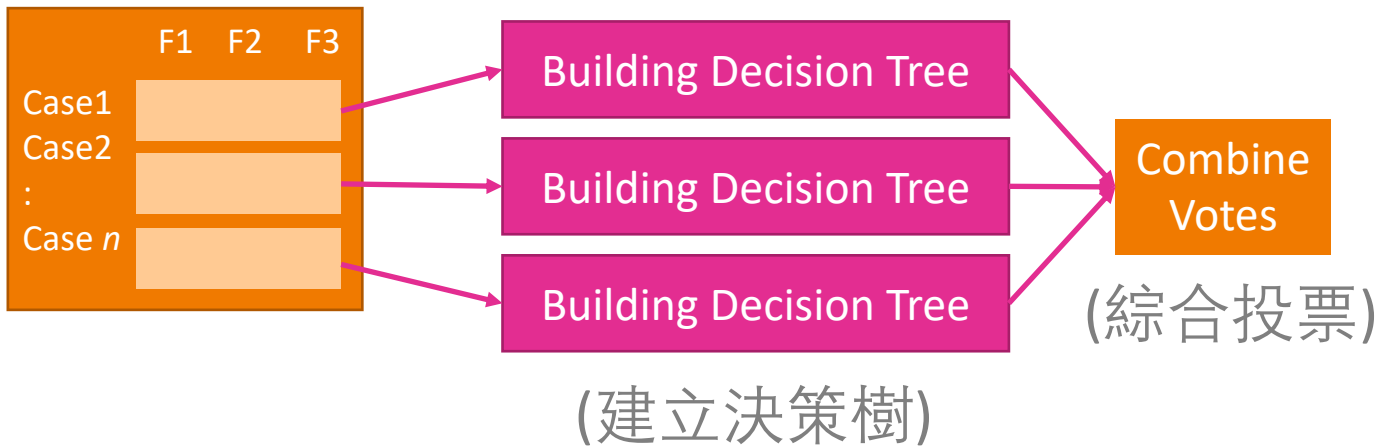
1. 不可能找出最佳的樹（太難了）
2. 樹常常會長得太複雜（overfitting 過擬合）
3. training data (用來訓練的資料) 一點變動，樹的變化大



Bootstrap Aggregating 引導聚集算法 (Bagging 裝袋算法)

□ 一次訓練多個model (模型) 來預測的方法

Training Set (訓練組合)

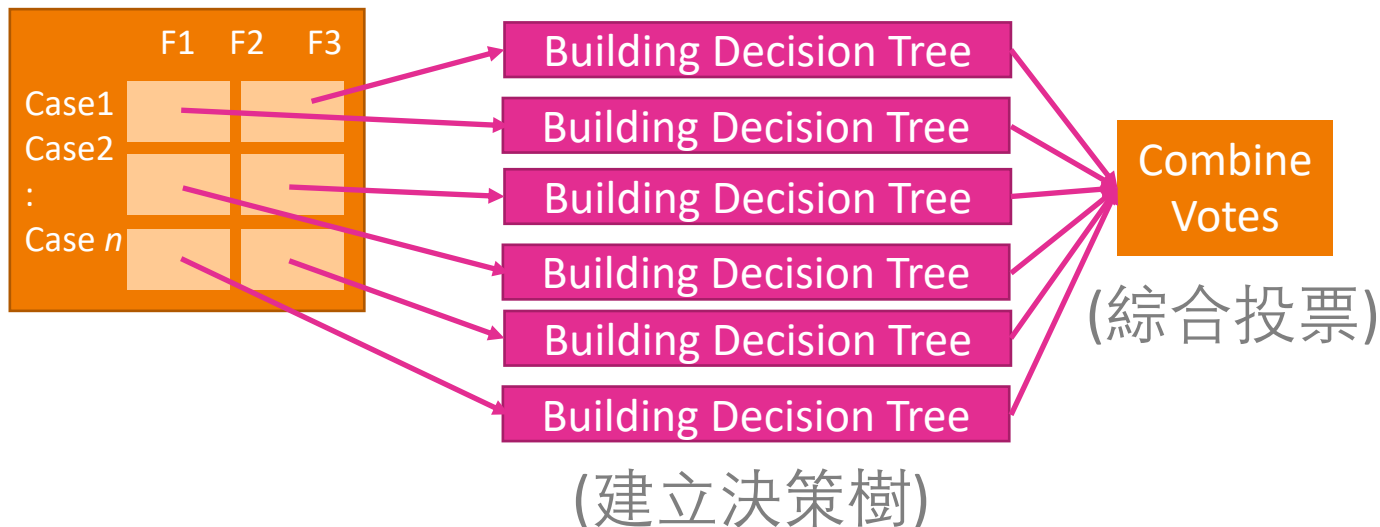




Random Forest Model (隨機森林模型)

- 隨機選擇部分 features (特徵) 來建構 decision tree (決策樹)

Training Set (訓練組合)





Random Forest Algorithm 隨機森林演算法總結

- ❑ Decision Tree (決策樹)
- ❑ ID3 Algorithm (疊代二元樹3代演算法)
 - ❑ Entropy (熵)
 - ❑ Information Gain (熵的變化量)
- ❑ Bagging 裝袋算法 (bootstrap aggregating 引導聚集算法)



謝謝您的收看

下個單元見！