

FDA_HW3-1 Report

F74066307 陳品修

1.資料選擇與前處理:

以 Open Price · High Price · Low Price 與 Volume 預測 Close Price

對 Volume 的處理是直接除 10 的 9 次方，讓它變成大於 0 小於 10 的 float 數

對 Open · High · low Price 做標準化

對 Close 也做標準化，不過 Label 要求要是整數，因此做完標準化之後乘以 10 之後捨去小數位(`astype(int)`)，保留兩位數值

2.三個模型預測:

三個都用 K-fold 拆 train set 做訓練

· 數值 acc 是直接用 `accuracy_score` 算，而漲跌 accuracy 是，用下一天的 close 剪今天的 close，如果 pred 跟 test 出來的走向一致的話（都是大於或小於或等於），給 acc+1，最後讓 acc 除掉總數算正確率

1.Logistic Regression:

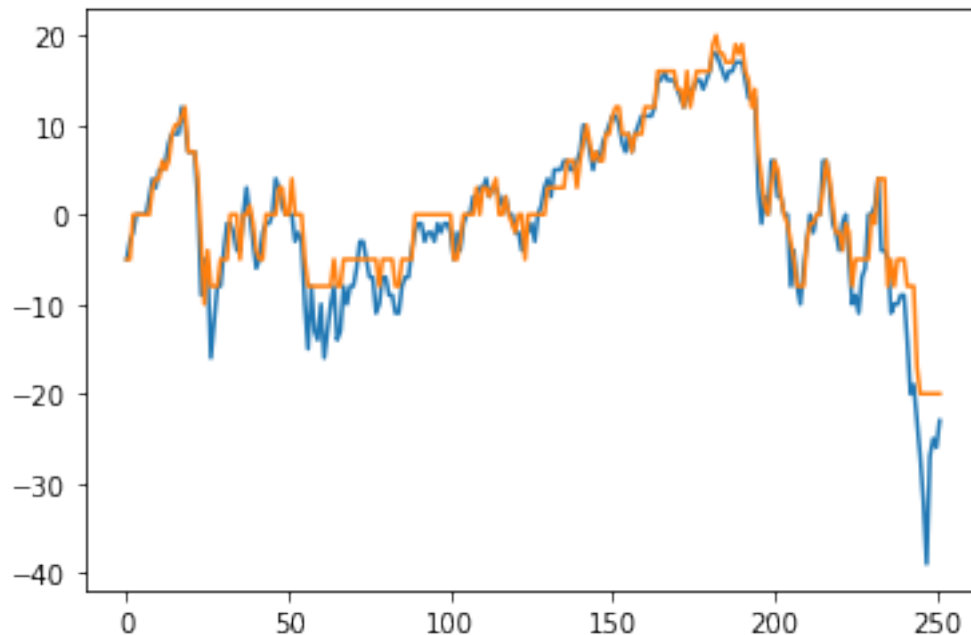
利用 `sklearn.linear_model` 的 `LogisticRegression`

```
model = LogisticRegression(random_state  
=1012,max_iter=1000,solver='saga',class_weight='balanced')
```

固定 random state，選擇 saga 作為 solver，在 `max_iter=1000` 附近能得到較好的結果，在 500 或 1500 時的效果都沒有 1000 的結果好，指定 `class_weight` 會比沒有指定效果好
結果:

數字 acc:0.21428571428571427 ,股價漲跌 accuracy: 0.44223107569721115

Close Price 圖:



2.neural network:

利用 sklearn.neural_network 的 MLPClassifier

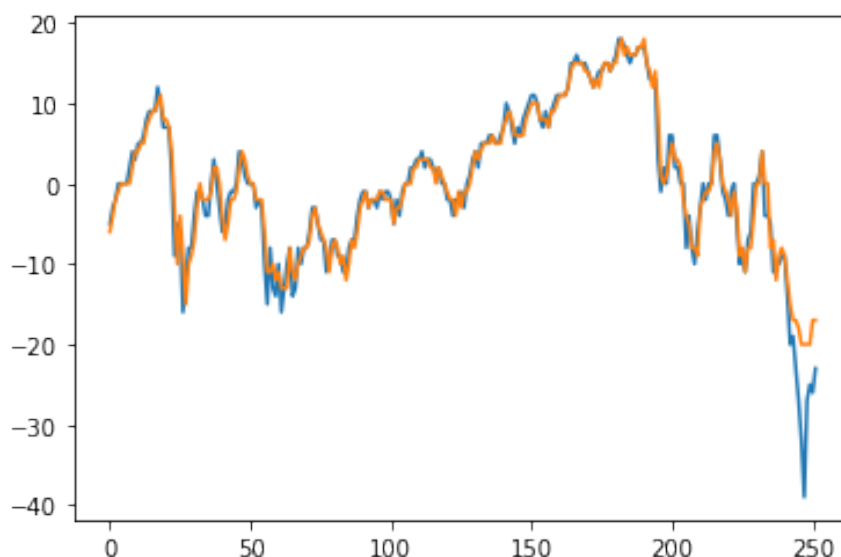
```
model = MLPClassifier(random_state=1012,max_iter=500,hidden_layer_sizes=(100,200,10))
```

固定 random state , hidden_layer_sizes 試了幾種 , 在這樣的情況會得到不錯的效果 , 第二層的數字比一三層多的情況結果貌似會較好

結果:

數字 acc:0.31746031746031744 , 股價漲跌 accuracy: 0.5577689243027888

Close price
圖:



3.decison tree classifier:

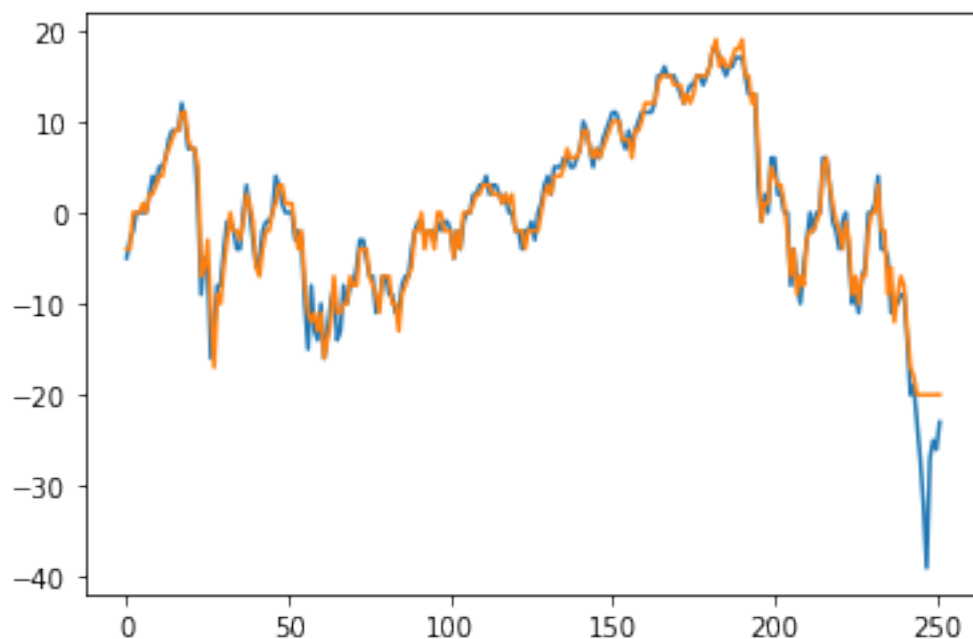
利用 sklearn.tree 的 DecisionTreeClassifier

```
model = DecisionTreeClassifier(  
    criterion='gini',  
    random_state=1012)
```

結果:

數字 acc:0.2857142857142857 , 股價漲跌 accuracy: 0.5816733067729084

預測圖:



3.結論:

決策樹預測走向效果比想像中的好，沒有特地調參數也能出現挺不錯的結果

Logistic Regression 還不是很了解運作，以及要如何調整，因此看來結果是最差的，而且會出現預測時無法收斂的問題

神經網路的預測圖看起來是最符合走向的，但是在 test 最後面一路降低的地方還是不夠符合，可能是訓練資料中較少出現這類情況，導致預測不到這種走向