

# 分類模型比較簡報

109703061 資料三 張翔

我從老師限定的五個分類模型中選擇三個來相互比較，並且將三個模型寫在同一份程式內。  
在我的程式中，我將訓練檔案 fashion-mnist\_train.csv 內的 6000 筆資料全數拿來做為訓練資料，模型訓練完畢後，用測試檔案 fashion-mnist\_test.csv 內的資料做預測，並且透過 `accuracy_score()` 和 `time.time()` 計算個別模型的預測正確率及訓練時間。

## ✓ 執行環境

CPU : Intel(R) Core(TM) i5-10300H CPU @ 2.50GHz

記憶體：16 GB

## ✓ 嘗試的分類模型 (Decision trees, Nearest neighbors, Naïve Bayes model)

- |  |  |   |
|--|--|---|
| • DecisionTreeClassifier() <ul style="list-style-type: none"><li>◦ 正確率：81.97 %</li><li>◦ 訓練時間：8.1927 s</li></ul> | • KNeighborsClassifier() <ul style="list-style-type: none"><li>◦ 正確率：85.89 %</li><li>◦ 訓練時間：0.1655 s</li></ul> | • MultinomialNB() <ul style="list-style-type: none"><li>◦ 正確率：66.75 %</li><li>◦ 訓練時間：0.0488 s</li></ul> |
|--|--|---|

```
Accuracy of DecisionTree_model : 0.8197
Training time of DecisionTree_model : 8.192688465118408 seconds
-----
Accuracy of NaïveBayes_model : 0.8589
Training time of NaïveBayes_model : 0.16553258895874023 seconds
-----
Accuracy of NearestNeighbors_model : 0.6675
Training time of NearestNeighbors_model : 0.04884672164916992 seconds
```

+ Code    + Markdown

## ✓ 調整分類模型參數

- |   |   |  |
|---|---|--|
| • DecisionTreeClassifier() <ul style="list-style-type: none"><li>◦ <code>random_state=10</code><br/>(default=None)</li><li>◦ <code>splitter="random"</code><br/>(default='best')</li><li>◦ <code>max_depth=13</code><br/>(default=None)</li></ul> | • KNeighborsClassifier() <ul style="list-style-type: none"><li>◦ <code>leaf_size=50</code><br/>(default=30)</li></ul> | • MultinomialNB() <ul style="list-style-type: none"><li>*無調整參數</li></ul> |
|---|---|--|

## ✓ 調整分類模型參數之意義

- |  |  |
|--|--|
| • DecisionTreeClassifier() <ul style="list-style-type: none"><li>◦ <code>random_state=10</code><br/>Decision tree 本身具有隨機性，固定該參數讓模型穩定(固定隨機性)</li><li>◦ <code>splitter="random"</code><br/>Decision tree 在分支時會更具隨機性以降低 overfitting 的可能性</li><li>◦ <code>max_depth=13</code><br/>限制 Decision tree 的深度，以避免 overfitting</li></ul> | • KNeighborsClassifier() <ul style="list-style-type: none"><li>◦ <code>leaf_size=50</code><br/>影響 tree 建構的速度以及搜索速度，但修改這個參數對於預測正確率並沒有顯著提升</li></ul> |
|--|--|