# 分類模型比較簡報

109703061 資科三 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()

KNeighborsClassifier()

MultinomialNB()

。 正確率:81.97%

。 正確率:85.89 %

。 正確率:66.75 %

。訓練時間:6.9958 s

。 訓練時間:0.0470 s

。 訓練時間:0.2062 s

Accuracy of DecisionTree model: 0.8197 Training time of DecisionTree\_model : 6.995762825012207 seconds Accuracy of NearestNeighbors\_model : 0.8589 Training time of NearestNeighbors\_model : 0.047075510025024414 seconds Accuracy of NaïveBayes\_model : 0.6675 Training time of NaïveBayes\_model : 0.20620989799499512 seconds + Code

+ Markdown

## ✔ 調整分類模型參數

- DecisionTreeClassifier()
- KNeighborsClassifier()
- MultinomialNB()

\*無調整參數

o random\_state=10 (default=None)

- o leaf\_size=50
- (default=30)
- o splitter="random" (default='best')
- o max\_depth=13 (default=None)

### ✔ 調整分類模型參數之意義

- DecisionTreeClassifier()
  - o random\_state=10

Decision tree 本身具有隨機性,固定該參數讓 模型穩定(固定隨機性)

o splitter="random"

Decision tree 在分支時會更具隨機性以降低 overfitting 的可能性

o max\_depth=13

限制 Decision tree 的深度,以避免 overfitting

- KNeighborsClassifier()
  - o leaf\_size=50

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

分類模型比較簡報 1