

題目：隨機森林迴歸預測模型

隨機森林是經典的監督式學習算法之一，請透過該算法進行 `cars` 資料之迴歸數值預測。

- 相關套件提示
 - Python 套件：`pandas, from sklearn.ensemble import RandomForestRegressor`

10% 小題一：讀取資料

題目說明

- 請讀取 `data` 資料夾內 `cars.csv` 資料集

答案示意

- (Python) 前五筆資料的輸出應如下：

##	speed	dist
## 0	4	2
## 1	4	10
## 2	7	4
## 3	7	22
## 4	8	16

20% 小題二：切分訓練集與測試集

題目說明

- 請以資料筆數的 70% 和 30% 將 `cars` 資料隨機抽樣分為兩份，訓練集與測試集
 - 因具有隨機性質，抽樣結果可能不同，這部分不硬性規定，但若希望與答案示意結果相同可參考下方設定：
 - Python: `random_state=1` (此部分非必要執行)

答案示意

- (Python) 輸出應如下：

訓練集

##	speed	dist
## 27	16	40
## 35	19	36
## 40	20	52
## 38	20	32
## 2	7	4
## 3	7	22
## 48	24	120
## 29	17	40
## 46	24	92
## 31	18	42
## 32	18	56
## 39	20	48
## 21	14	60
## 36	19	46
## 19	14	26
## 42	20	64
## 49	25	85

```
## RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
##                        max_features='auto', max_leaf_nodes=None,
##                        min_impurity_decrease=0.0, min_impurity_split=None,
##                        min_samples_leaf=1, min_samples_split=2,
##                        min_weight_fraction_leaf=0.0, n_estimators=10,
##                        n_jobs=None, oob_score=False, random_state=None,
##                        verbose=0, warm_start=False)
```

20% 小題四：預測

題目說明

- 以配適後模型進行測試集資料之預測
 - 放入模型的應為測試集資料的 speed 欄位，並得到相同數量的預測結果

答案示意

- (Python) 預測結果應如下：
 - 因為抽樣與訓練過程是隨機，**數值僅供參考**，主要會看過程是否正確

```
## RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
##                          max_features='auto', max_leaf_nodes=None,
##                          min_impurity_decrease=0.0, min_impurity_split=None,
##                          min_samples_leaf=1, min_samples_split=2,
##                          min_weight_fraction_leaf=0.0, n_estimators=10,
##                          n_jobs=None, oob_score=False, random_state=None,
##                          verbose=0, warm_start=False)
##
## [18.25      18.25      17.2      26.8      26.8      26.8
##  27.6      25.93333333 25.93333333 37.53333333 37.53333333 28.21666667
##  41.68571429 52.3      86.93666667]
```

20% 小題五：評估

題目說明

- 計算每個預測結果與實際結果之誤差，因為誤差可能有正有負，請將其各別平方後平均，即 MSE(Mean Squared Error)。

答案示意

- (Python) 評估結果如下：
 - 因為抽樣與訓練過程是隨機，**數值僅供參考**，主要會看過程是否正確

```
## RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
##                          max_features='auto', max_leaf_nodes=None,
##                          min_impurity_decrease=0.0, min_impurity_split=None,
##                          min_samples_leaf=1, min_samples_split=2,
##                          min_weight_fraction_leaf=0.0, n_estimators=10,
##                          n_jobs=None, oob_score=False, random_state=None,
##                          verbose=0, warm_start=False)
##
## 120.30138595616027
```