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
      7
## 2
          4
      7 22
## 3
## 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
## 49 25
          85
```

```
## 26
         16
               32
## 22
         14
                80
## 13
         12
               24
## 41
         20
               56
## 17
         13
               34
## 45
         24
               70
## 24
         15
               26
## 23
         15
               20
## 4
          8
               16
## 33
         18
               76
## 14
         12
               28
## 30
         17
               50
## 10
         11
               28
## 28
         17
               32
## 44
         23
               54
## 34
         18
               84
## 18
         13
               46
               36
## 20
         14
## 測試集
##
       speed dist
## 0
          4
                2
## 1
          4
               10
## 5
          9
               10
## 6
         10
               18
## 7
         10
               26
## 8
         10
               34
## 9
         11
               17
## 11
         12
               14
         12
## 12
               20
         13
## 15
               26
## 16
         13
               34
## 25
         15
               54
## 37
         19
               68
## 43
         22
               66
## 47
         24
               93
```

30% 小題三:模型配適

題目說明

- 使用訓練集進行隨機森林模型配適
 - speed 為自變數(x), dist 為應變數(y)

答案示意

- (Python) 配適後模型的輸出應如下:
 - 僅僅示意於 Python 中直接將模型輸出的結果,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)
```

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,<mark>數值僅供參考</mark>,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
                                       26.8
                                                   26.8
                                                                26.8
               18.25
                           17.2
               25.9333333 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,<mark>數值僅供參考</mark>,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
```