

✧ Random Forest Regression

- ✧ **方法簡介：**整體學習 (Ensemble learning) 可以將數個分類器(決策樹)的預測結果綜合考慮，藉此達到顯著提升分類的效果。

✧ 建置步驟

- step1 : pick at random K data point from Training set(N)(Bagging)
從樣本數(N)中，決定抽取 K 項 data()
- step2 : build decision tree(Build Tree)
建造決策數
- step3 : choose the number Ntree of tree(repeat step1 and step2)
重複 step1 & step2(記得每次建完 tree，要將 K 項 data()放回 N)，並依照需求決定建幾棵樹
- step4 : to vote and decide the result(Ensemble)
採多數決，決定結果

✧ 載入資料

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset = pd.read_csv('Position_Salaries.csv')
X = dataset.iloc[:, 1:2].values
y = dataset.iloc[:, 2].values
```

✧ 建立預測模型

```
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators = 10, random_state = 0)
regressor.fit(X, y)
```

✧ RandomForestRegressor 說明

n_estimators:決定建造幾棵樹，太低或太高會使預測結果位移，基本值為 100 來做測試(大多從 100 開始試)

random_state:確保每次執行結果為相同

✧ 進行預測

執行 `y_pred = regressor.predict(6.5)`

(1) `n_estimators = 10` · 預測結果為 167000

(2) `n_estimators = 100` · 預測結果為 158300

(3) `n_estimators = 300` · 預測結果為 160333.333

(4) `n_estimators = 500` · 預測結果為 160600

(5) `n_estimators = 1000` · 預測結果為 161600

✧ 補充

隨機森林其實就是由很多個決策樹組合而成的，那如果把 `n_estimators` 設成 1 的預測結果是不是會和決策樹的預測結果相同？

決策樹的預測結果:150000

隨機森林(`n_estimators=1`)的預測結果：200000

其實會有這樣的差異主要是這 2 種方法背後運算的方式其實不太一樣，決策樹是將所有的樣本(N)一起丟下去跑，而隨機森林是會從所有樣本(N)中隨機抽取 K 的觀測值來跑，跑完再將 K 放回(N)，所以產生這樣的差異。