

Day 19

隨機森林 (迴歸器)

[全民瘋AI系列]



第12屆 iT邦幫忙 鐵人賽

Day 19 學習目標

01

隨機森林 (迴歸)

了解隨機森林是如何處理連續性數值輸出

02

實作隨機森林迴歸器

觀察隨機森林中的樹的數量是否影響預測

Part 1

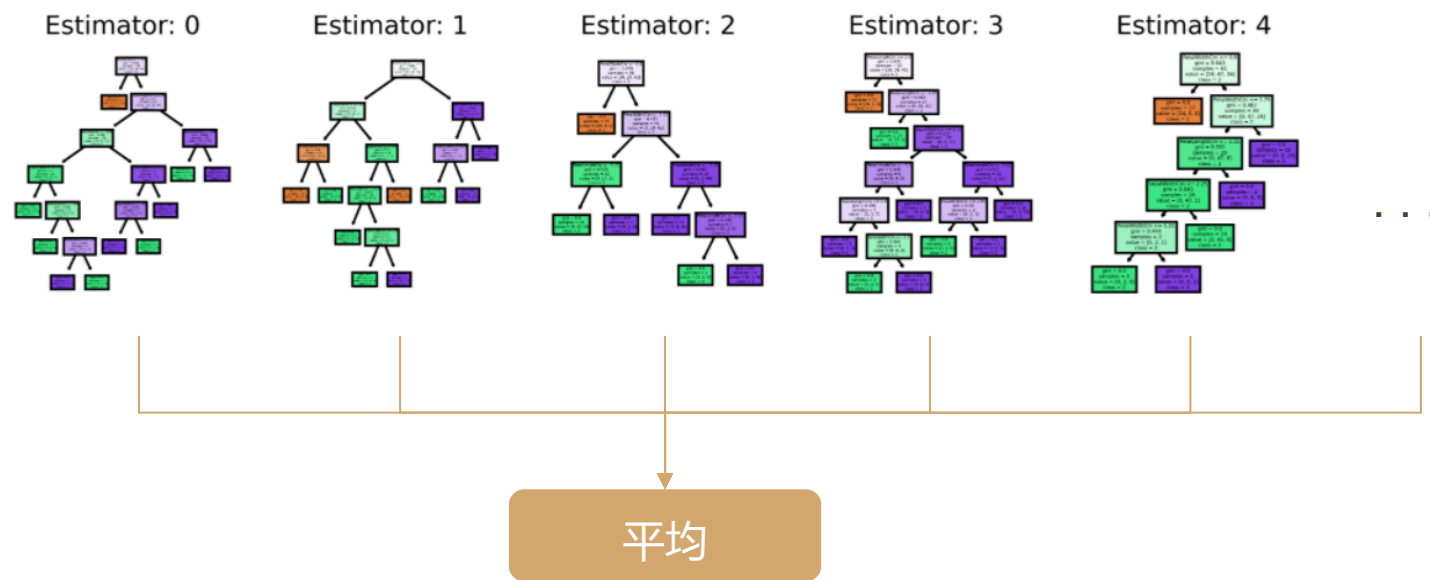
隨機森林 (迴歸器) 觀念講解



第12屆 iT邦幫忙 鐵人賽

隨機森林(迴歸器)的生成方法

1. 從訓練集中抽取 n' 筆資料出來
2. n' 筆資料隨機挑選 k 個特徵做樣本
3. 重複 m 次，產生 m 棵決策樹
4. 平均機制進行預測



//// 隨機森林中的隨機？

隨機森林中的隨機有兩種方面可以解釋：

1. **隨機取樣：** 從訓練集中抽取 n' 筆資料出來，這 n' 筆資料是可以被重複抽取的。
2. **隨機特徵選取：** n' 筆資料隨機挑選 k 個特徵做樣本。

Bootstrap

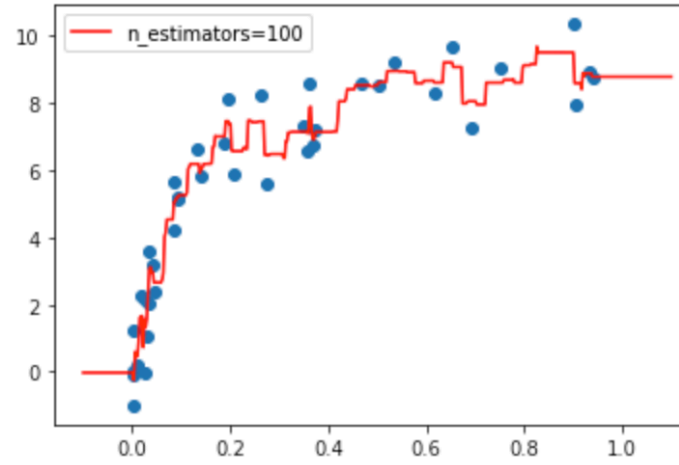
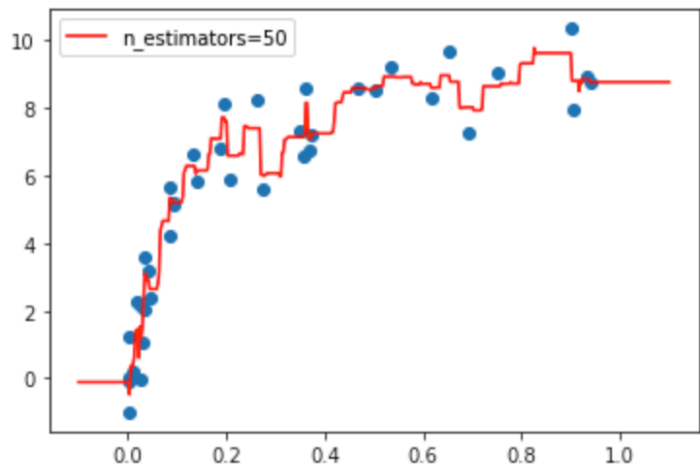
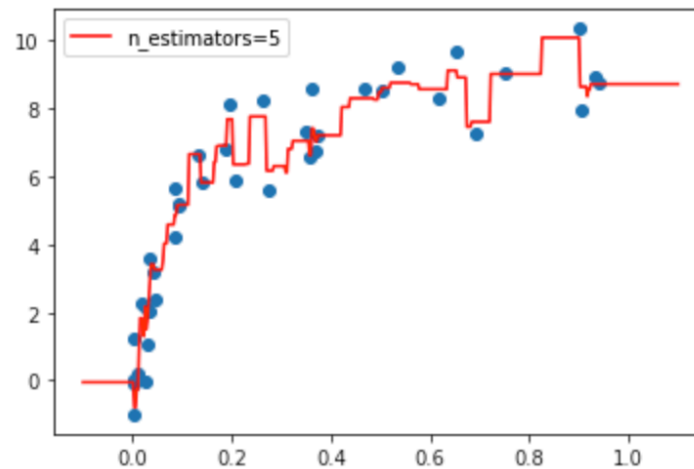
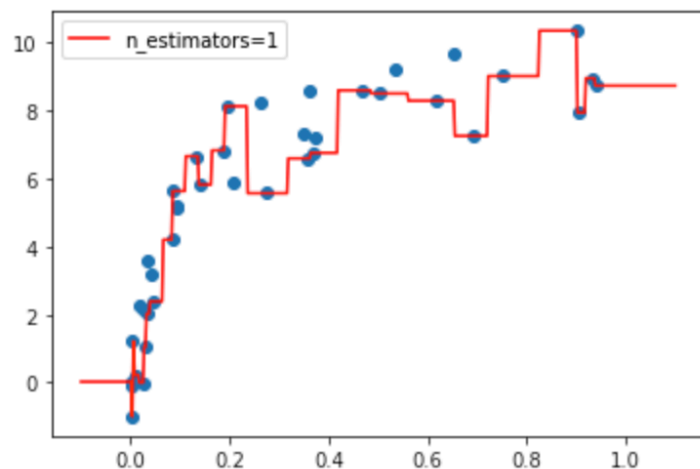


在 sklearn 中，最多隨機選取 $\log_2 N$ 個特徵



樹越多模型越複雜

$n_estimators$ 太小，容易欠擬合， $n_estimators$ 太大，計算量會變久。



迴歸隨機森林該如何選擇切割點？

透過評估指標找出誤差最小的值作為樹的特徵選擇與切割點

- Mean Square Error
- Mean Absolute Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$



Part 2

隨機森林 (迴歸器) 程式實作



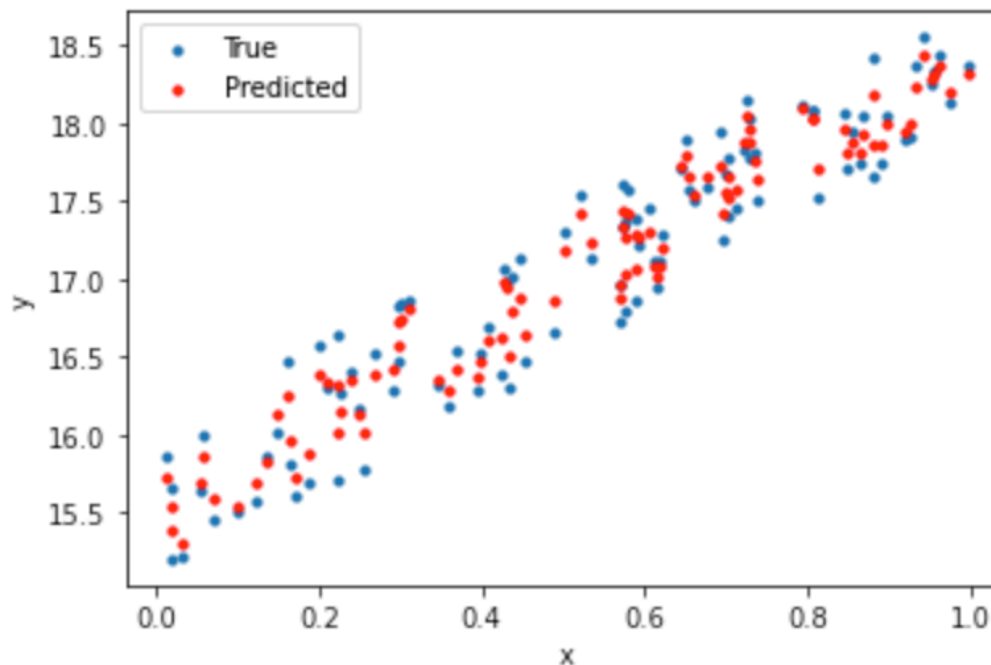
第12屆 iT邦幫忙 鐵人賽

Random Forest Regressor 簡單線性迴歸

```
from sklearn.ensemble import RandomForestRegressor

# 建立RandomForestRegressor模型
randomForestModel = RandomForestRegressor(n_estimators=100, criterion = 'mse')
# 使用訓練資料訓練模型
randomForestModel.fit(x, y)
# 使用訓練資料預測
predicted=randomForestModel.predict(x)
```

- 目標函式: $y=3x+15$
- 隨機添加 noise 讓資料分散
- x 值域介於 0~1



非線性迴歸

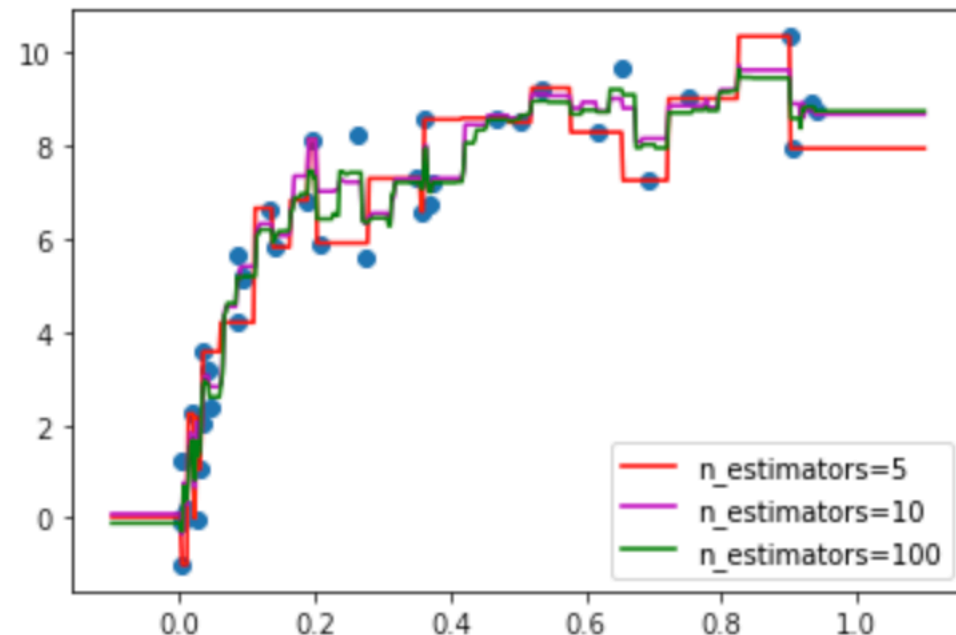
動手試試看調整不同的 `n_estimators` 是否能夠提升預測。

Parameters:

- `criterion`: 評估切割點指標, `mse/friedman_mse/mae`
- `n_estimators`: 森林中樹木的數量
- `max_features`: 劃分時考慮的最大特徵數, 預設`auto`



觀察隨機森林中的樹的數量是否影響預測



Thanks

PRESENTED BY 10程式中



第12屆 iT邦幫忙 鐵人賽