

D44 : tree based model - 隨機森林程式碼撰寫



簡報閱讀



範例與作業



問題討論

隨機森林 - 程式碼撰寫



知識地圖



本日知識點目標



使用 Sklearn 中的隨機森林



隨機森林的模型超參數



延伸閱讀



解題時間



隨機森林 - 程式碼撰寫



知識地圖

機器學習基礎模型建立

監督式學習
Supervised Learning



非監督式學習
Unsupervised Learning



模型選擇 Model selection

概論	基礎模型 Basic Model	樹狀模型 Tree based Model
驗證基礎	線性回歸 Linear Regression	決策樹 Decision Tree
預測類型	邏輯斯回歸 Logistic Regression	隨機森林 Random Forest
評估指標	套索算法 LASSO	梯度提升機 Gradient Boosting Machine
	嶺回歸 Ridge Regression	

本日知識點目標

今日知識點目標

- 了解隨機森林的程式碼應用
- 如何用 Scikit-learn 撰寫隨機森林的程式碼
- 使用內建的特徵重要性進行特徵選取

使用 Sklearn 中的隨機森林

如同決策樹的使用方式，根據不同問題 import 不同的模型

可以看到是從 sklearn.ensemble 這裏 import 的，代表隨機森林是個**集成**模型，透過多棵複雜的決策樹來投票得到結果，緩解原本決策樹容易過擬和的問題，實務上的結果通常都會比決策樹來得好

```
from sklearn.ensemble import  
RandomForestClassifier  
from sklearn.ensemble import  
RandomForestRegressor  
clf = RandomForestRegressor()
```

隨機森林的模型超參數

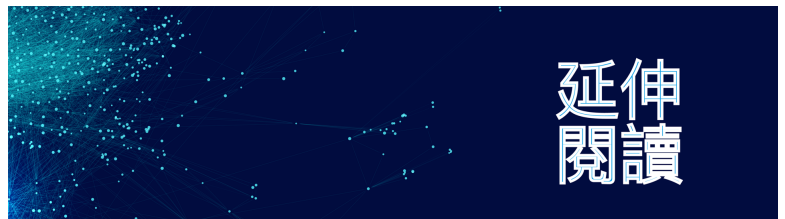
- 同樣是樹的模型，所以像是 max_depth, min_samples_split 都與決策樹相同
- 可決定要生成數的數量，越多越不容易過擬和，但是運算時間會變長

```
from sklearn.ensemble import  
RandomForestClassifier
```

```
clf = RandomForestClassifier()
```

```
max_features="auto", #如何選取 features
max_depth=10,
min_samples_split=2,
min_samples_leaf=1
)
```

延伸閱讀

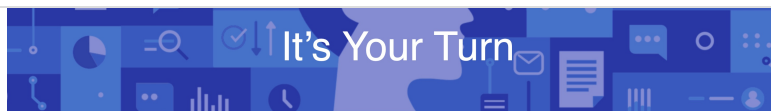


- 知名 ML youtuber 教你手刻隨機森林 by Python

Random Forests - The Math of Intellig...



解題時間



Sample Code & 作業
開始解題



[下一步：閱讀範例與完成作業](#)

