

Day 44

機器學習

隨機森林 - 程式碼撰寫



出題教練

楊証琨

知識地圖 機器學習- 模型選擇 - 隨機森林 程式碼撰寫



機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning



非監督式學習 Unsupervised Learning



模型選擇 Model selection

概論

驗證基礎

預測類型

評估指標

基礎模型 Basic Model

線性回歸 Linear Regression

邏輯斯回歸 Logistic Regression

套索算法 LASSO

嶺回歸 Ridge Regression

樹狀模型 Tree based Model

決策樹 Decision Tree

隨機森林 Random Forest

梯度提升機 Gradient Boosting Machine

本日知識點目標

- 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(  
    n_estimators=10, #決策樹的數量  
    criterion="gini",  
    max_features="auto", #如何選取 features  
    max_depth=10,  
    min_samples_split=2,  
    min_samples_leaf=1  
)
```

解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業
開始解題

