

Day 42

機器學習

# 決策樹 - 程式碼撰寫



Coding 練習日



出題教練

楊証琨



# 知識地圖 機器學習- 模型選擇 - 決策樹 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

- 了解決策樹的程式碼應用
- 如何用 Scikit-learn 撰寫決策樹的程式碼
- 使用內建的特徵重要性進行特徵選取

# 使用 Sklearn 建立決策樹模型

---

根據回歸/分類問題分別建立不同的 Classifier

- `from sklearn.tree_model import DecisionTreeRegressor`
- `from sklearn.tree_model import DecisionTreeClassifier`
- `clf = DecisionTreeClassifier()`

# 決策樹的超參數

- Criterion: 衡量資料相似程度的 metric
- Max\_depth: 樹能生長的最深限制
- Min\_samples\_split: 至少要多少樣本以上才進行切分
- Min\_samples\_leaf: 最終的葉子 (節點) 上至少要有多少樣本

```
from sklearn.tree import DecisionTreeClassifier

clf = DecisionTreeClassifier(
    criterion = 'gini',
    max_depth = None,
    min_samples_split = 2,
    min_samples_leaf = 1,
)

# feature importance
clf.feature_importances_
```

# 解題時間 It's Your Turn

請跳出PDF至官網Sample Code & 作業  
開始解題

