Day 20

XGBoost (分類器)

[全民瘋AI系列]

第12屆 iT邦幫忙 鐵人賽

Day 20 學習目標

01

XGBoost 介紹

XGBoost 是什麼？為什麼它那麼強大？

02

Bagging vs. Boosting

比較兩種集成式學習架構差異

03

實作 XGBoost 分類器

比較 Bagging 與 Boosting 兩者差別

Part 1

XGBoost (分類器)

觀念講解

第12屆 iT邦幫忙 鐵人賽

XGBoost

XGBoost 是由華盛頓大學博士班學生陳天奇所開發，是目前 Kaggle 競賽中最常見到的算法。

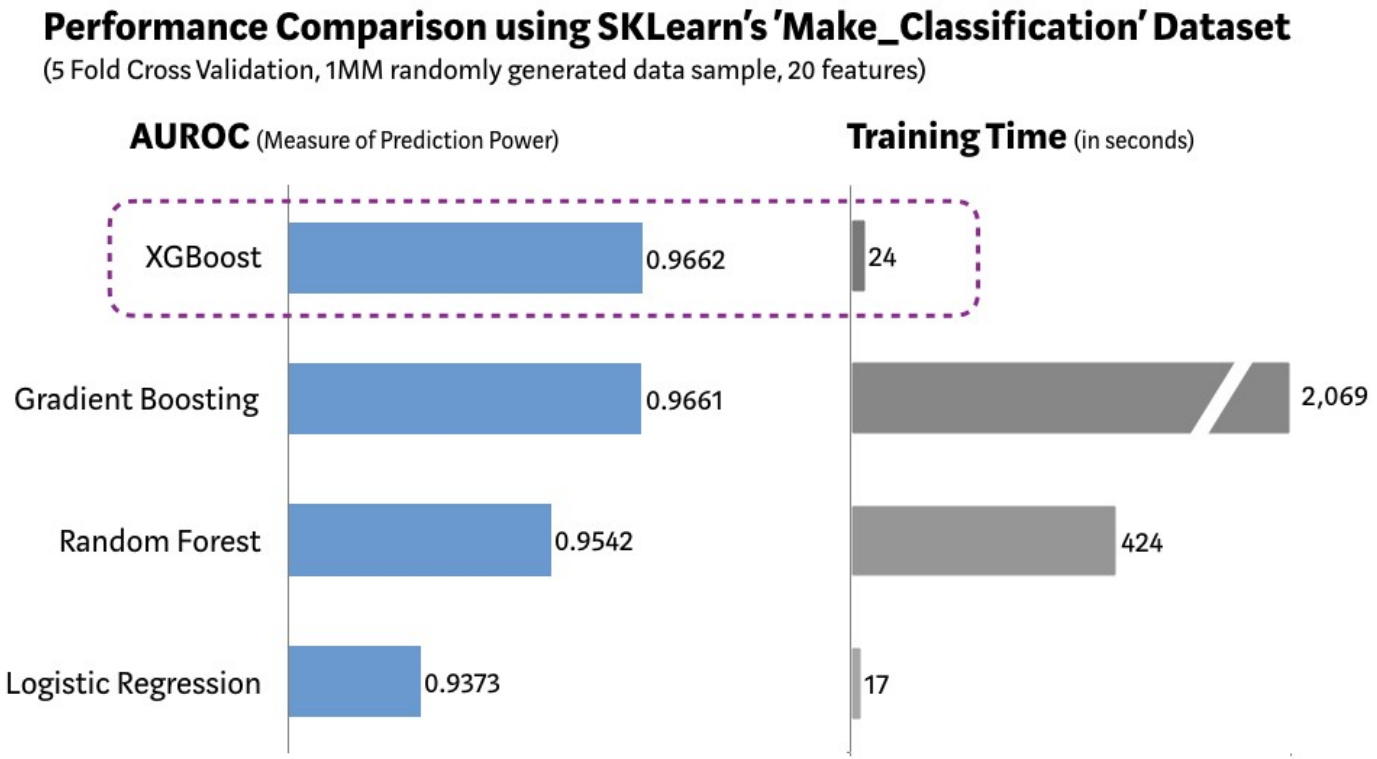
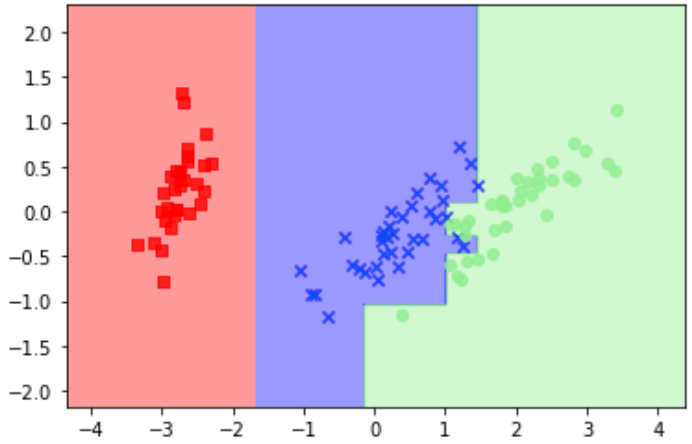


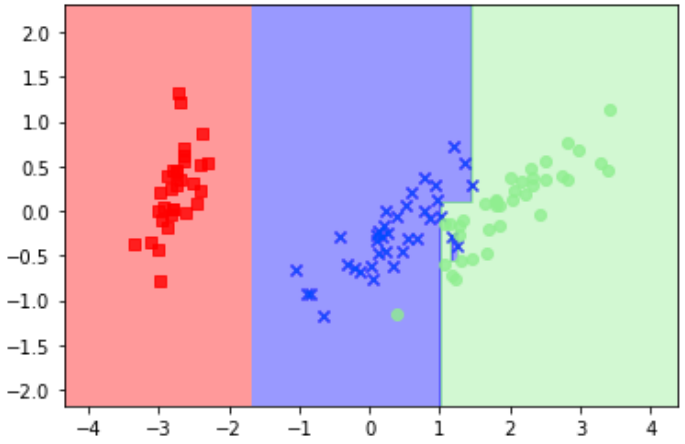
Image from: towards data science

XGBoost

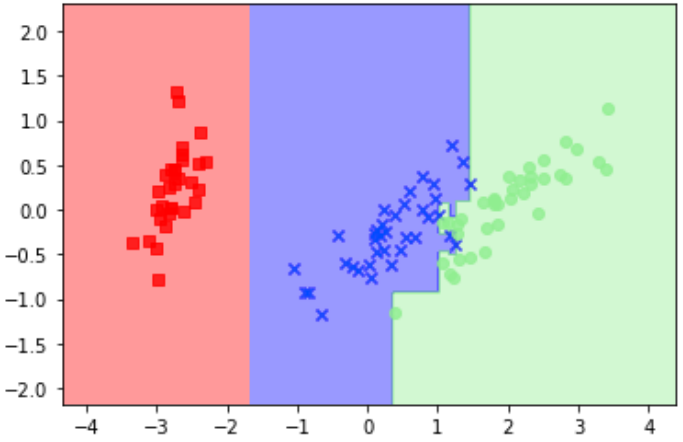
- XGBoost 全名為eXtreme Gradient Boosting
- 以 Gradient Boosting 為基礎下去實作
- 每一棵樹是互相關聯的
- 和隨機森林一樣採用特徵隨機採樣的技巧
- 是 Ensemble learning 中的 Boosting 的實例



決策樹



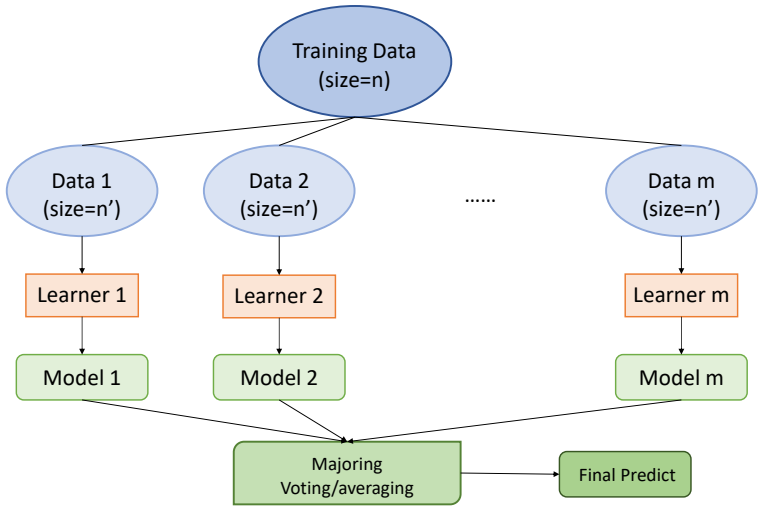
Boosting with 5 estimators



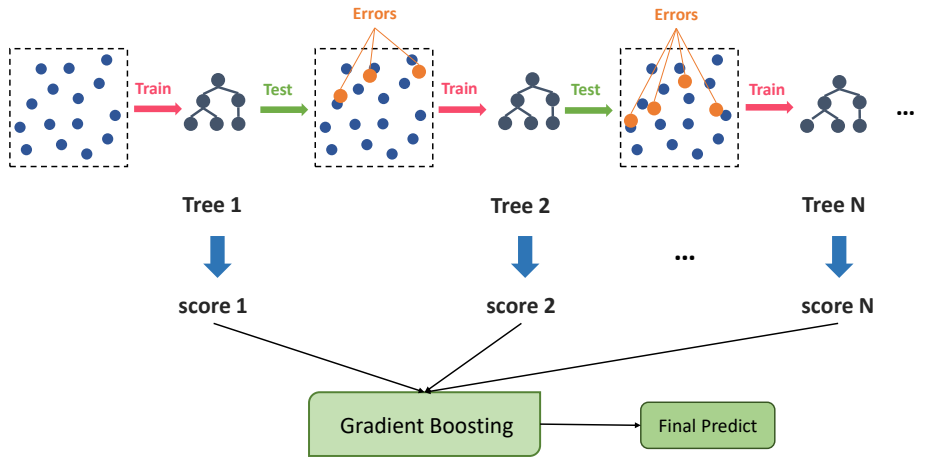
Boosting with 100 estimators

Bagging vs. Boosting

- **Bagging** 透過抽樣的方式生成樹，每棵樹彼此獨立
- **Boosting** 透過序列的方式生成樹，後面生成的樹會與前一棵樹相關



Bagging



Boosting

試著用 XGBoost 訓練

Example : 鳶尾花朵

Part 2

XGBoost (分類器)

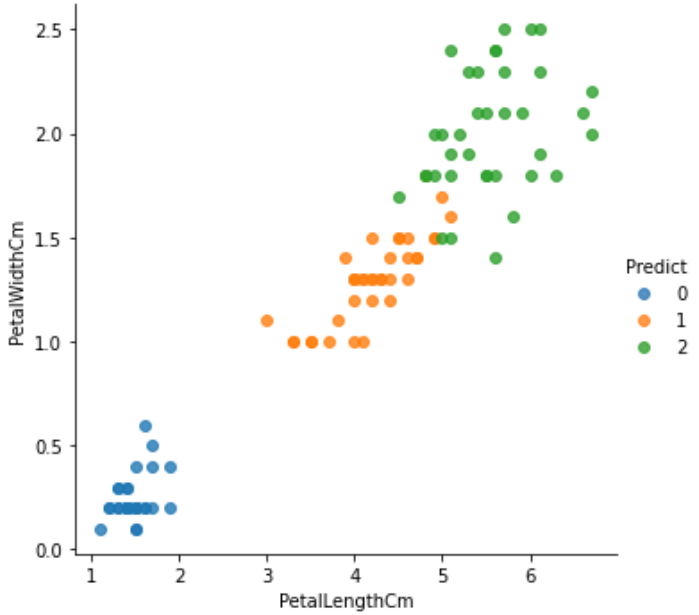
程式實作

XGBoost Classifier

```
from xgboost import XGBClassifier

# 建立XGBClassifier模型
xgboostModel = XGBClassifier(n_estimators=100, learning_rate= 0.3)
# 使用訓練資料訓練模型
xgboostModel.fit(X_train, y_train)
# 使用訓練資料預測分類
predicted = xgboostModel.predict(X_train)
```

隨機森林 (訓練集)預測結果

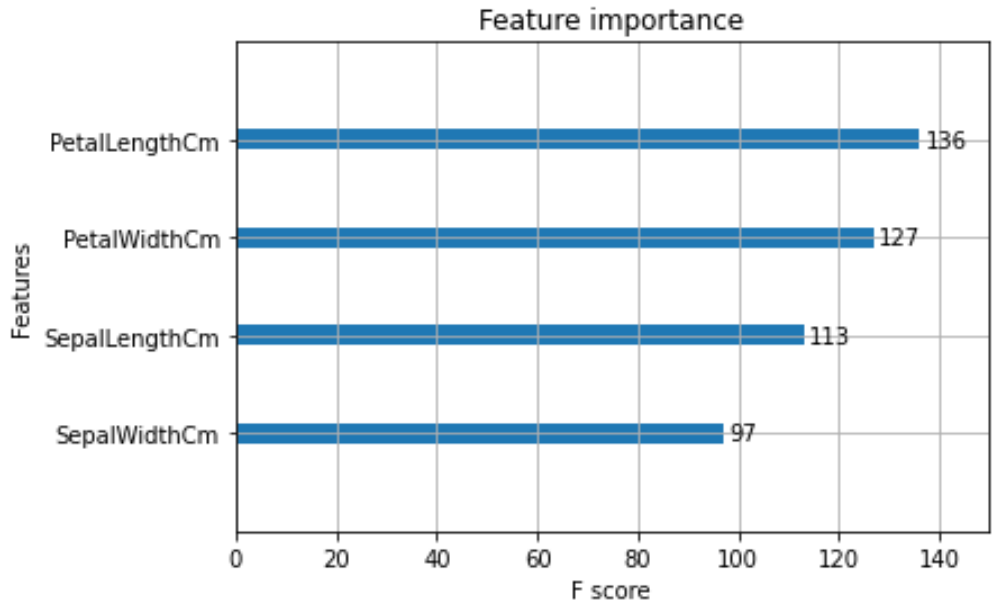


特徵重要程度

```
from xgboost import plot_importance
from xgboost import plot_tree

plot_importance(xgboostModel)
print('特徵重要程度: ',xgboostModel.feature_importances_)
```

```
特徵重要程度:  [0.01001516 0.03135139 0.7407739  0.21785954]
```



Thanks

PRESENTED BY 10程式中

第12屆 iT邦幫忙 鐵人賽