# Day 36 機器學習

# 評估指標選定

出題教練 楊証琨

# 知識地圖 機器學習- 模型選擇 - 評估指標選定(Evaluation metrics)

CUPOY

## 機器學習概論 Introduction of Machine Learning

**監督式學習**
**Supervised Learning**

前處理 Processing → 探索式數據分析 Exploratory Data Analysis → 特徵工程 Feature Engineering → **模型選擇 Model selection** → 參數調整 Fine-tuning → 集成 Ensemble

**非監督式學習**
**Unsupervised Learning**

- 分群 Clustering
- 降維 Dimension Reduction

### 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 Logistic Regression
- 梯度提升機 Gradient Boosting Machine

# 本日知識點目標

- 了解機器學習中評估指標的意義及如何選取
- 迴歸、分類問題應選用的評估指標
- 不同評估指標的意義及何時該使用

# 評估指標

CUPOY

- 設定各項指標來評估模型預測的準確性，最常見的為準確率 (Accuracy) = **正確分類樣本數/總樣本數**

- 不同評估指標有不同的評估準則與面向，衡量的重點有所不同

# 評估指標 - 回歸

CUPOY

觀察「**預測值**」 (Prediction) 與「**實際值**」 (Ground truth) 的差距

- MAE, Mean Absolute Error, 範圍: [0, ∞]
- MSE, Mean Square Error, 範圍:  [0, ∞]
- R-square, 範圍:  [0, 1]

# 評估指標 - 分類

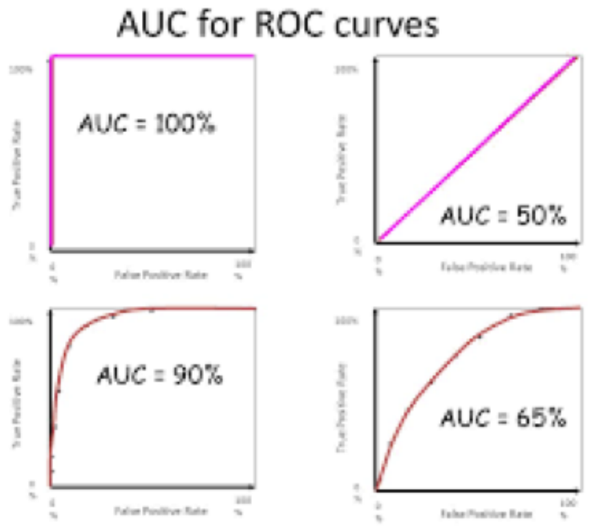
觀察「**預測值**」 (prediction) 與「**實際值**」 (Ground truth) 的正確程度

- AUC, Area Under Curve, 範圍: [0, 1]
- F1 - Score (Precision, Recall), 範圍: [0, 1]

# 評估指標 - 分類 - AUC, Area Under Curve

CUPOY

- AUC 指標是分類問題常用的指標，通常分類問題都需要定一個閾值 (threshold) 來決定分類的類別 (通常為機率 > 0.5 判定為 1，機率 < 0.5 判定為 0)
- AUC 是衡量曲線下的面積，因此可考量所有閾值下的準確性，因此 AUC 也廣泛地在分類問題的比賽中使用




圖片來源：slidesplayer

# 評估指標 - 分類 - F1-Score

CUPOY

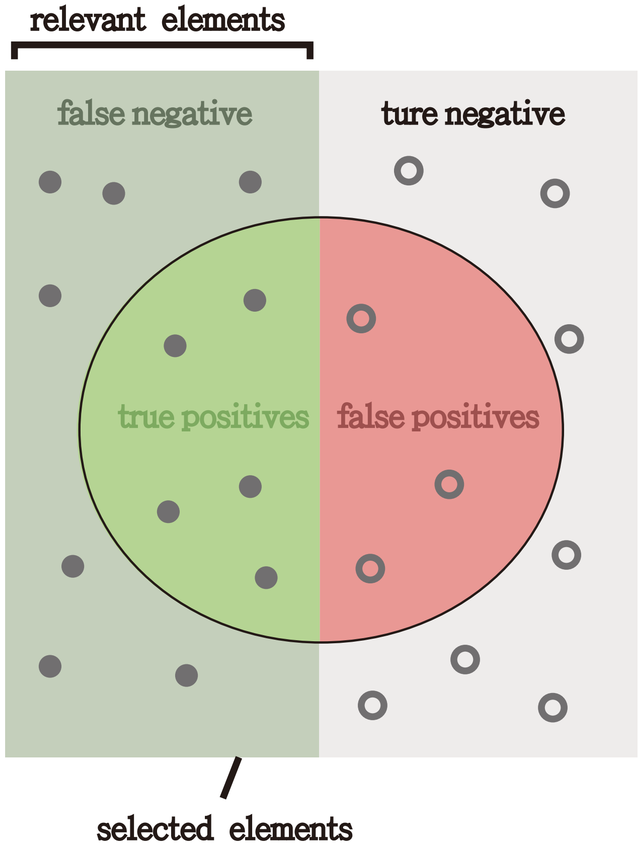
- 分類問題中，我們有時會對某一類別的準確率特別有興趣。例如瑕疵/正常樣本分類，我們希望任何瑕疵樣本都不能被漏掉。
- Precision，Recall 則是針對某類別進行評估
  - Precision: 模型判定瑕疵，樣本確實為瑕疵的比例
  - Recall: 模型判定的瑕疵，佔樣本所有瑕疵的比例

    (以瑕疵檢測為例，若為 recall=1 則代表所有瑕疵都被找到)
- F1-Score 則是 Precision, Recall 的調和平均數

# 一張圖理解 Precision, Recall

CUPOY

- 右圖可看到 Precision 與 Recall 的公式
- 其中有四個值，True Positive, False Positive, True Negative, False Negative
- T,F 代表模型預測對或錯，P/N 代表模型預測結果
- 例如 True Positive 代表模型預測是正樣本且預測正確！




How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

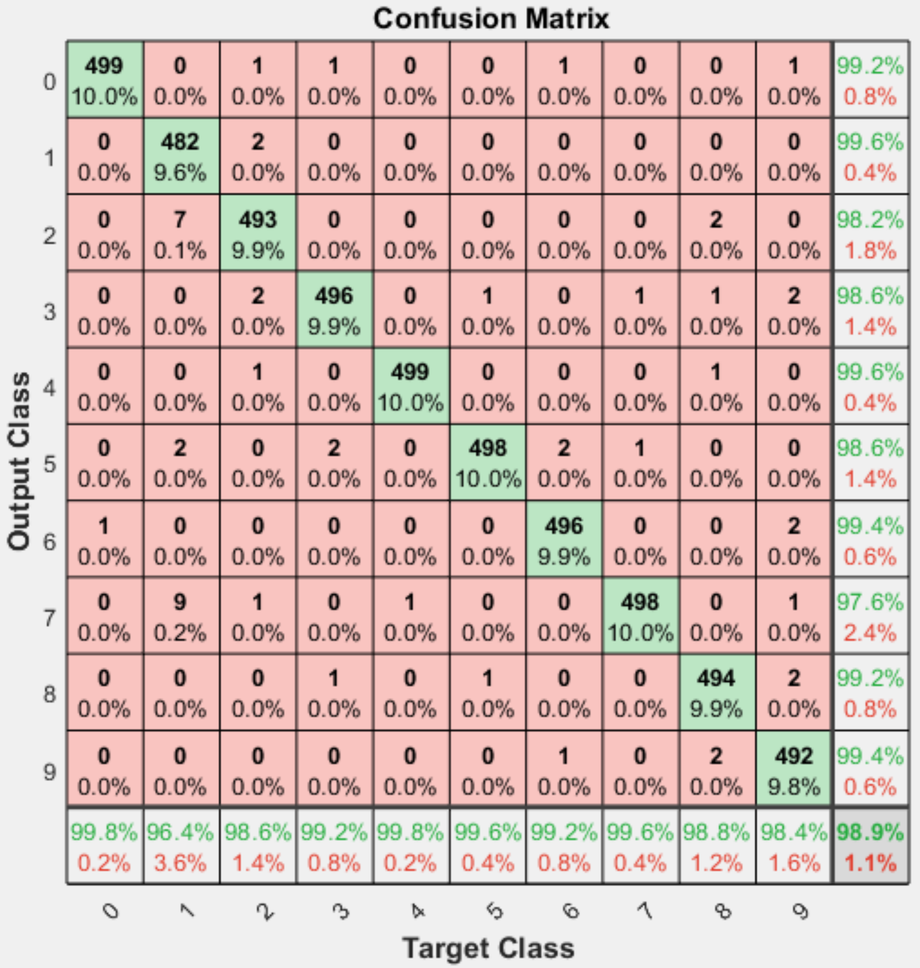
Precision：被分類器挑選(selected)出來的正體樣本究竟有多少是真正的樣本
Recall：在全部真正的樣本裡面分類器選了多少個！

圖片來源：saowen

# 評估指標 - 分類 - 混淆矩陣 (Confusion Matrix)

CUPOY

- 縱軸為模型預測
- 橫軸為正確答案
- 可以清楚看出每個 Class 間預測的準確率，完美的模型就會在對角線上呈現 100 % 的準確率




圖片來源：MathWorks

# 常見問題

Q：這麼多評估指標，該怎麼選擇?

A：**回歸問題**可以透過 R-square 很快了解預測的準確程度；**分類問題**若為二分類 (binary classification)，通常使用 AUC 評估。但如果有特別希望哪一類別不要分錯，則可使用 F1-Score，觀察 Recall 值或是 Precision 值。若是多分類問題，則可使用 top-k accuracy，k 代表模型預測前 k 個類別有包含正確類別即為正確 (ImageNet 競賽通常都是比 Top-5 Accuracy)

# 常見問題

Q：Sklearn 的 AUC 計算結果怪怪的？F1-Score 計算時出現錯誤？

A：AUC 計算時 y_pred 的值必須填入每個樣本的預測機率 (probability) 而非分類結果！

A：F1-Score 計算時則需填入每個樣本已分類的結果，如機率 >=0.5 則視為 1，而非填入機率值

# 解題時間
# It’s Your Turn

請跳出PDF至官網Sample Code＆作業
開始解題