

# 機器學習

## 效能衡量指標

國立屏東科技大學資訊管理系

吳庭育 教授

[tyw@mail.npust.edu.tw](mailto:tyw@mail.npust.edu.tw)

# 效能衡量指標

- 混淆矩陣(Confusion Matrix)
- 準確率(Accuracy)
- 精確率(Precision)
- 召回率(Recall)
- F1 Score
- 真陽率(True Positive Rate)
- 假陽率(False Positive Rate)
- ROC/AUC 曲線

# 混淆矩陣(Confusion Matrix)及相關效能衡量指標

		True condition			
Total population		Condition positive	Condition negative	Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$	Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$
Predicted condition	Predicted condition positive	<b>True positive</b> , Power	<b>False positive</b> , Type I error	Positive predictive value (PPV), Precision = $\frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$
	Predicted condition negative	<b>False negative</b> , Type II error	<b>True negative</b>	False omission rate (FOR) = $\frac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection = $\frac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\sum \text{False positive}}{\sum \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR}^+}{\text{LR}^-}$  F <sub>1</sub> score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$
		False negative rate (FNR), Miss rate = $\frac{\sum \text{False negative}}{\sum \text{Condition positive}}$	True negative rate (TNR), Specificity (SPC) = $\frac{\sum \text{True negative}}{\sum \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

Confusion matrix -- Wiki

# 混淆矩陣(Confusion Matrix)

- TP (True Positive) — 真陽：
  - 即實際為True，預測為Positive。預測的結果與實際情況相同。
- TN (True Negative) — 真陰：
  - 即實際為True，預測為Negative。預測的結果與實際情況相同。
- FP (False Positive) — 偽陽：
  - 即實際為False，預測為Positive。預測的結果與實際情況不同。
- FN (False Negative) — 偽陰：
  - 即實際為False，預測為Negative。預測的結果與實際情況不同。

	「模型預測」為真 (positive)	「模型預測」為非 (negative)
「真實情況」為真	true positive (TP)	false negative (FN)
「真實情況」為非	false positive (FP)	true negative (TN)

# 混淆矩陣範例

- 有一模型用來預測165名病患是否患有癌症(Cancer)其混淆矩陣如下

	模型預測，Cancer = Yes	模型預測，Cancer = No
真實情況，Cancer = Yes	true positive (TP) = 20	false negative (FN) = 25
真實情況，Cancer = No	false positive (FP) = 20	true negative (TN) = 100

- 病患確實患有癌症且預測該患者患有癌症的人數為20人(TP)
- 病患並無患有癌症且預測該患者無患有癌症的人數為100人(TN)
- 病患並無患有癌症且預測該患者患有癌症的人數為20人(FN)
- 病患確實患有癌症且預測該患者無患有癌症的人數為25人(FN)

但光看這些數字，似乎無法很直覺的判別這個模型的好壞!!

準確率(Accuracy)、精確率(Precision)  
、召回率(Recall)、F1-source



# 準確率(Accuracy)

- 準確率(Accuracy) =  $(TP+TN)/(TP+FP+FN+TN)$  全部資料總數
  - 模型預測正確的比例
  - 在正向例子很少的狀況下，此種指標不適用。

$$\text{Accuracy(準確率)} = \frac{TP+TN}{\text{全部資料總數}}$$

Model正確預測的機率值

在正向例子很少的狀況下，此種指標不適用

	「模型預測」為真 (positive)	「模型預測」為非 (negative)
「真實情況」為真	true positive (TP)	false negative (FN)
「真實情況」為非	false positive (FP)	true negative (TN)

# 精確率(Precision)

- 精確率(Precision) =  $TP / (TP + FP)$ 
  - 即陽性的樣本中有幾個是預測正確

$$\text{Precision(精確率)} = \frac{TP}{TP + FP}$$

預測正向的狀況中，正確預測的機率值  
較在意「**預測正向**」的答對數量

	「模型預測」為真 (positive)	「模型預測」為非 (negative)
「真實情況」為真	true positive (TP)	false negative (FN)
「真實情況」為非	false positive (FP)	true negative (TN)



# 召回率(Recall)

- 召回率(Recall) =  $TP / (TP + FN)$ 
  - 即事實為真的樣本中有幾個是預測正確

$$\text{Recall(召回率)} = \frac{TP}{TP + FN}$$

實際正向的狀況中，實際為真的機率值  
較在意「實際正向」的答對數量

	「模型預測」為真 (positive)	「模型預測」為非 (negative)
「真實情況」為真	true positive (TP)	false negative (FN)
「真實情況」為非	false positive (FP)	true negative (TN)

# 召回率(Recall)

- 召回率是在所有正樣本當中，能夠預測多少正樣本的比例，準確率為在所有預測為正樣本中，有多少為正樣本。
- 舉例
  - 在小明家的門禁系統中，哪個比較重要呢？
    - 準確率應比較重要。
    - 我們希望判定成正樣本就一定要是正確的，不要有小王的臉可以打開小明家的門的情況。
    - 而召回率低的話，也不過是常常無法判斷出來小明的臉，但至少不是誤判。

# 精確率(Precision) VS 召回率(Recall)

- 一個Precision高而Recall低的模型跟一個Recall高而Precision低的模型，分別代表什麼意義呢？
- 前者，可以看作一個比較謹慎的模型，雖然常常沒辦法抓出命名實體，但只要有抓出幾乎都是正確的(Precision高)。
- 後者則是一個寬鬆的模型，雖然有時候會抓錯，但幾乎該抓的都有抓到(Recall高)。

# F1-Source

- **F1-Score** :  $F1 = 2 / ( (1/ \text{Precision}) + (1/ \text{Recall}) )$ 
  - 是**精確率(Precision)**和**召回率(Recall)**的調和平均數 ( Harmonic Mean )，可看作是該二指標的**綜合指標**，能較全面地評斷模型的表現。
- 在F1-Score中，會選用「調和平均數」而非「算數平均數」作為平均**精確率(Precision)**和**召回率(Recall)**的方式，是為了要強調**較小值**的重要性。
- 因為該二指標是互相制約的——不會有兩個指標同時很高或很低的狀況發生，這樣可以更加方便評價模型的好壞。

# F1-Source

當recall接近1、precision接近0

採用「調和平均數」的F1-score接近0 → 等效於評價precision和recall的整體效果

(代表模型的效果還有很大很大的進步空間)

$$F_{\beta} = (1 + \beta^2) \times \frac{\text{precision} \times \text{recall}}{(\beta^2 \times \text{precision}) + \text{recall}}$$

- F1-source ( $\beta = 1$ ) : Precision和Recall同等重要
- F2-source ( $\beta = 2$ ) : Recall比Precision重要
- F0.5-source ( $\beta = 0.5$ ) : Precision比Recall重要

- 例子

- 根據統計美國每年有1000萬人次出入境，假設有10個恐怖份子試圖闖關，海關有抓到5個嫌疑犯。
- 準確率(Accuracy) = 99.9999..%
  - $(1000\text{萬} - (10-5))/1000\text{萬} = 99.9999..%$
- 美國政府會滿意嗎？
  - 事實上有一半的恐怖份子沒抓到
- 如果改用召回率(Recall)計算
  - $5/10 = 50\%$ (就合理多)



# 小結

- 準確率(Accuracy)

- 正確預測的樣本數佔總預測樣本數的比值。
- 它不考慮預測的樣本是正例還是負例。**考慮全部樣本。**

- 精確率(Precision)

- 正確預測的正樣本數佔所有預測為正樣本的數量的比值。
- 也就是說所有預測為正樣本的樣本中有多少是真正的正樣本。**只關注預測為正樣本的部份。**

- 召回率(Recall)

- 正確預測的正樣本數佔真實正樣本總數的比值。
- 也就是從這些樣本中能夠**正確找出多少個正樣本。**

- F-score

- 相當於精確率(Precision)和召回率(Recall)的調和平均
- 召回率(Recall)和精確率(Precision)**任何一個數值減小，F-score都會減小，反之，亦然。**



# ROC/AUC 曲線

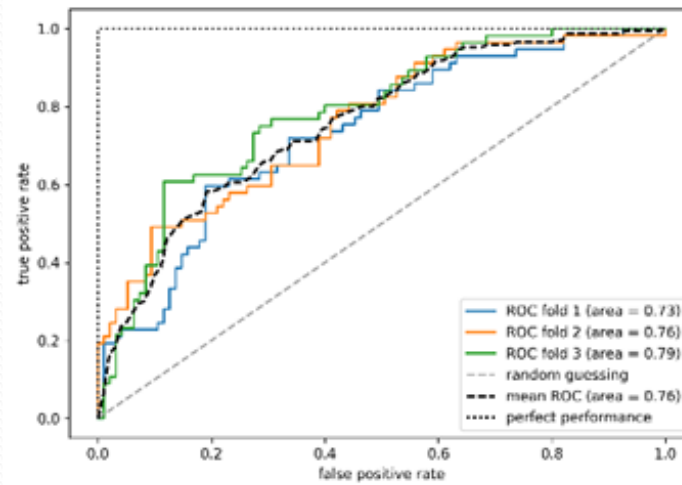
- ROC Curves (Receiver Operating Characteristic Curve)
  - 適合類別平均的情況，而 PR Curves 適合於類別不平均的情況
  - ROC 曲線以FPR為 X 軸，TPR為 Y 軸，每一個點代表設定不同的門檻值所得到的不同的 FPR及TPR，最後繪製成一條曲線。
  - 偽陽性率：FPR(False Positive Rate)
    - FPR表示成  $1 - \text{特異度}$  而特異度(Specificity)意指正確判斷出負樣本，故特異度越高、FPR越低，模型越能夠正確判斷負樣本、表現越好。
  - 真陽性率：TPR (True Positive Rate)
    - TPR又稱為敏感度(Sensitivity)，它也是我們熟知的召回率(Recall)，也就是正確判斷出正樣本，故TPR越高則模型越能夠正確判斷正樣本、表現越好
- AUC(Area under curve):曲線下面積
  - 可以當作一個判斷整體模型能力(Skill)的指標

# ROC/AUC 曲線

- ROC(Receiver Operator Characteristic) 接收操作特徵圖
  - 在各種『決策門檻』(Decision Threshold)下，比較『真陽率』(True Positive Rate; TPR)與『假陽率』(False Positive Rate; FPR)間的變化

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP + FN} \quad (\text{真陽率}) \\ \text{or Sensitivity} & \\ \text{Specificity} &= \frac{TN}{TN + FP} \quad (\text{假陽率})\end{aligned}$$

- ROC曲線可以繪製成一條曲線，如下圖，有多條ROC曲線，相互比較效能，AUC(Area Under the Curve)就比較容易理解，即ROC曲線之下所覆蓋的面積，除以總面積的比率。

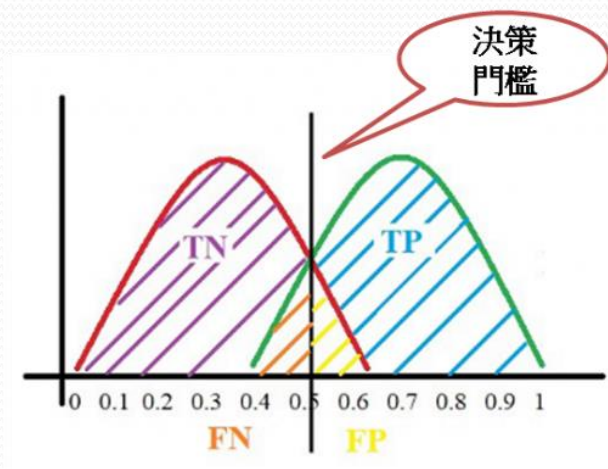


- 可以在各種門檻下計算真陽率及假陽率，作為樣本點，將所有樣本點連成一線，即ROC曲線。
- 因此，這條線越接近上方，表示真陽率越高，即判斷正確的比率越高，換句話說，ROC曲線下方覆蓋的面積(AUC)越大，表示效能越好。

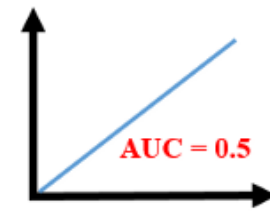
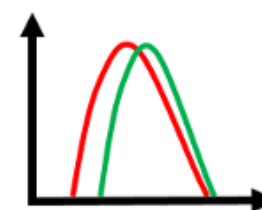
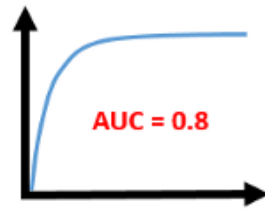
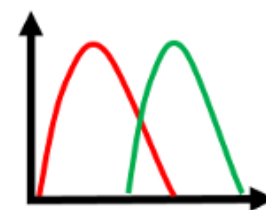
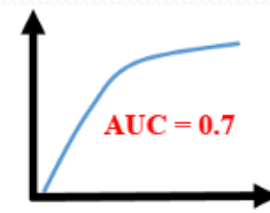
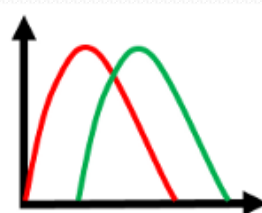
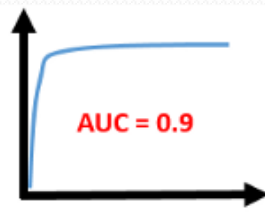
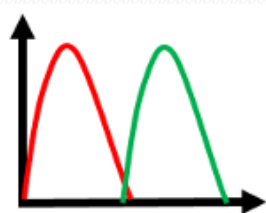
# AUC 鑑別度

- 利用AUC來概略地衡量模型整體表現
  - $AUC = 0.5$  (no discrimination 無鑑別力)
  - $0.7 \leq AUC \leq 0.8$  (acceptable discrimination 可接受的鑑別力)
  - $0.8 \leq AUC \leq 0.9$  (excellent discrimination 優良的鑑別力)
  - $0.9 \leq AUC \leq 1.0$  (outstanding discrimination 極佳的鑑別力)





		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN





END !!