

Homework : Receiver Operating Characteristic, ROC

III C51502, CY Chingyao Fu, NTUT

Oct 2023

Introduction

ROC (Receiver Operating Characteristic, 接收者操作特性) 曲線是一種用於評估二元分類模型性能的圖形工具。它是一個以偽正率 (False Positive Rate, FPR) 為橫軸, 真正率 (True Positive Rate, TPR) 為縱軸的圖。

I 主要概念

- 真正率 (True Positive Rate, TPR) : 也稱為敏感度 (Sensitivity), 是真正例 (True Positives, TP) 佔所有實際正例 (實際正例和假負例的總和) 的比例。

$$TPR = \frac{TP}{TP + FN}$$

- 偽正率 (False Positive Rate, FPR) : 是假正例 (False Positives, FP) 佔所有實際負例 (實際負例和真負例的總和) 的比例。

$$FPR = \frac{FP}{FP + TN}$$

2 ROC特點

- 對角線 : 對角線表示隨機猜測的情況, 此時模型沒有區分能力。
- 曲線下面積 (AUC) : ROC曲線下的面積被稱為AUC (Area Under Curve)。AUC值為1表示完美分類器, 而0.5表示無法區分的分類器。
- 閾值選擇 : ROC曲線可以幫助我們選擇一個適當的閾值, 該閾值能夠平衡真正率和偽正率。
- 不受類別不平衡影響 : ROC曲線的一個優點是它不受類別不平衡的影響。

3 實際應用

- 模型評估 : 通過比較不同模型的ROC曲線和AUC值, 我們可以選擇最佳的模型。
- 閾值調整 : 根據ROC曲線, 我們可以選擇一個適當的閾值, 以達到特定應用場景 (如醫療診斷、信用評分等) 的需求。
- 性能度量 : 除了AUC之外, ROC曲線也可以用來計算其他性能度量, 如Youden指數, 這是真正率和偽正率的差值。