

# 分類器的機率切點<sup>◎</sup>

國立東華大學電機工程學系 楊哲旻

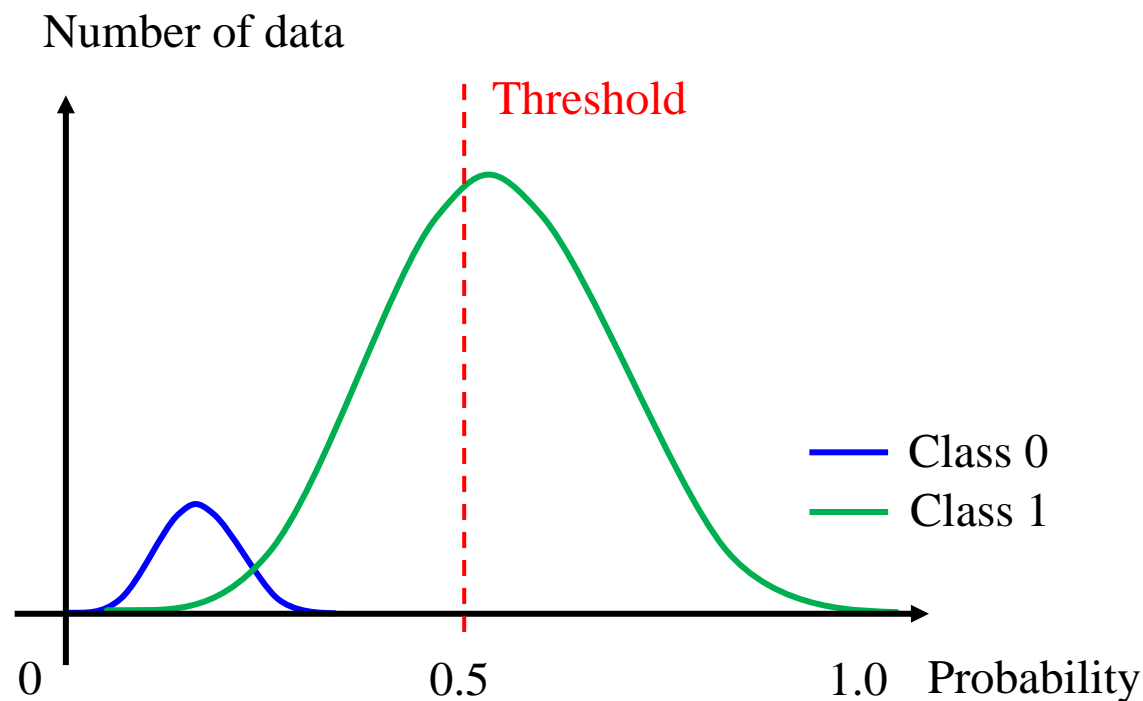
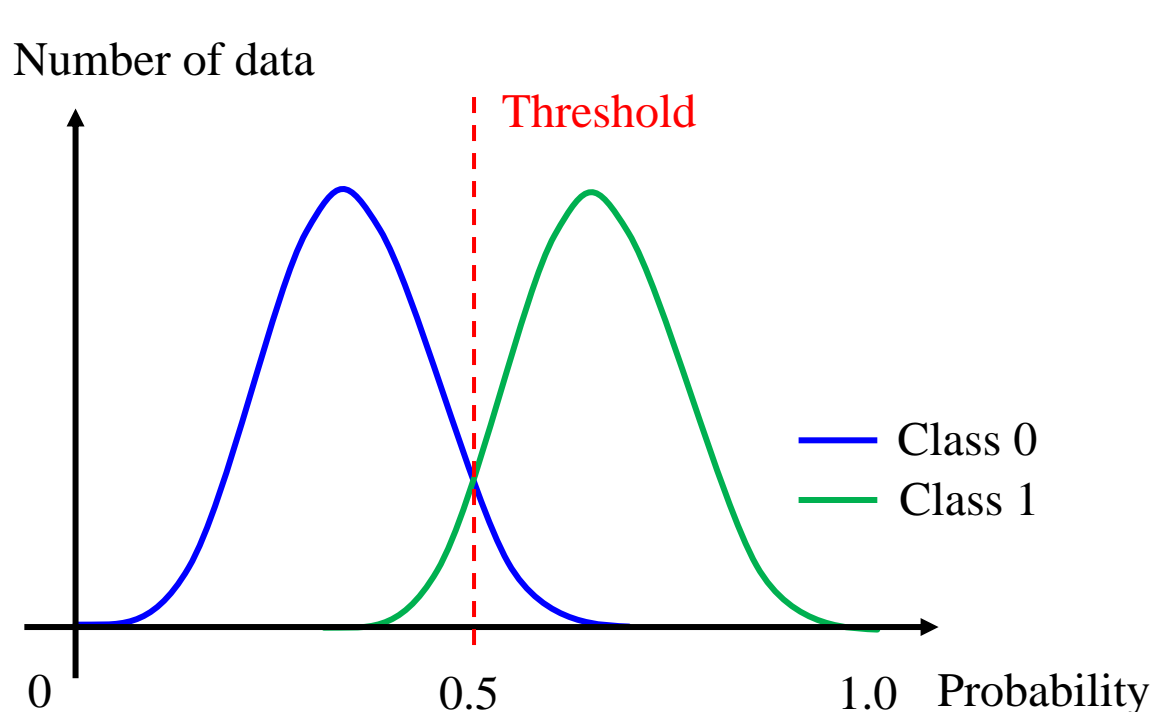
# Outline



- 1 最佳臨界值
- 2 指標的選擇
- 3 醫學檢測觀點
- 4 分類器的機率切點實作

### 最佳臨界值(Threshold)

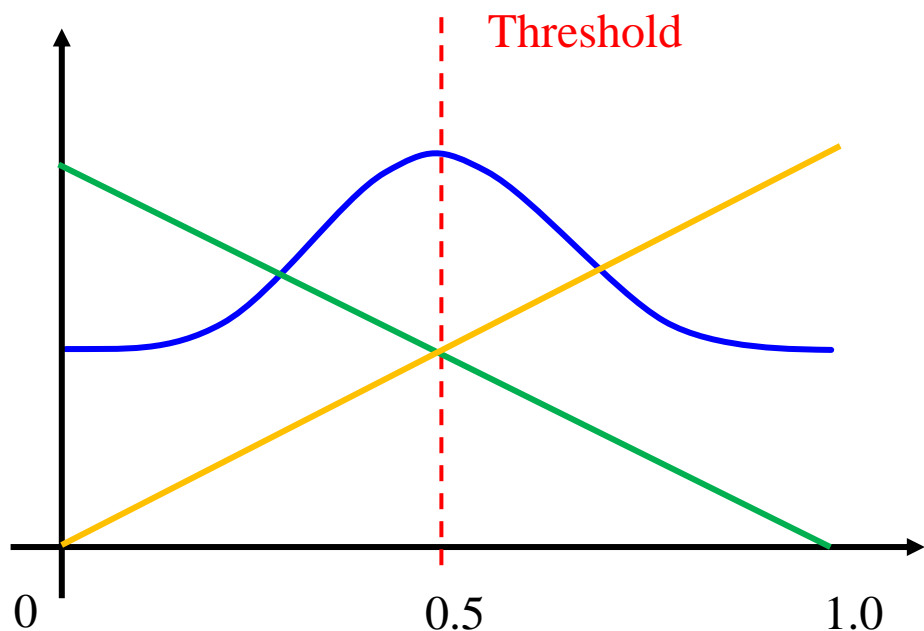
二元分類模型預測時回傳通常為一機率值，通常在均勻類別的常態分佈的資料中，我們預設會以0.5的臨界值來當作預測的標準，但在不均勻類別的資料中以0.5的臨界值可能預測效果不佳，如下兩圖



### 指標的選擇：均勻類別 v.s. 不均類別

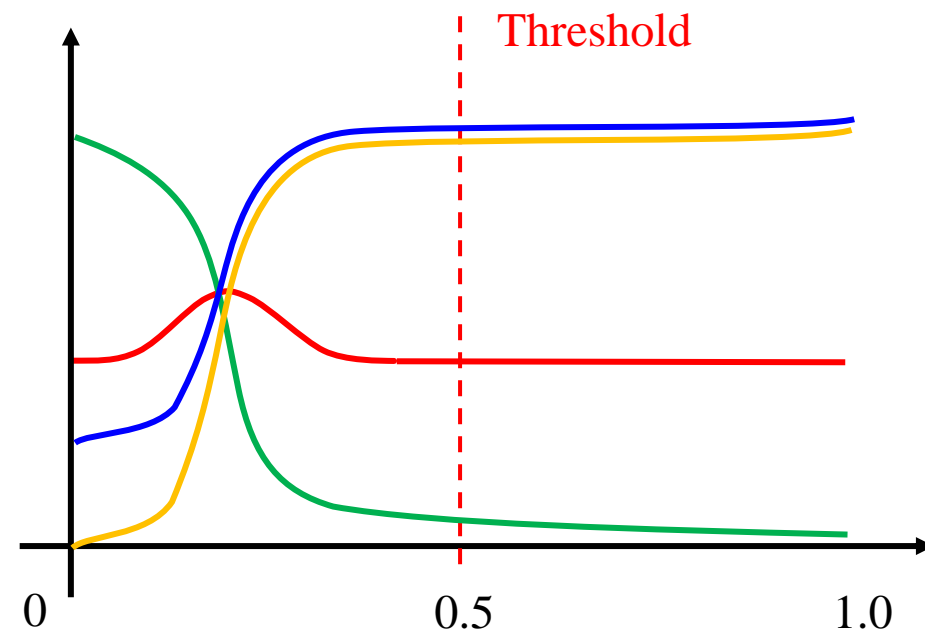
在均勻類別資料中，若繪製機率在所有臨界值下的常見指標，如「準確度、敏感度、特異度及F1分數」與機率臨界值繪製的軸線圖，會在臨界值0.5左右呈現對稱性，如左下圖；反之，在不均勻類別資料中，則呈非對稱性且有包含多數類別為分子的指標，如準確度、特異度等，容易在越過一機率值快速下滑，如右下圖

Number of data



— Accuracy,  $F_1$   
— Sensitivity  
— Specificity  
—  $F_1$

Number of data



# 分類器的機率切點

## 03.醫學檢測觀點

### 指標的選擇：均勻類別 v.s. 不均類別

在不均勻類別資料中，建議尋找最佳超參數時所採用指標以不均勻類別特為顯著的指標，如平衡準確度 (Balanced Accuracy), F1 分數 (F1 score) 等，並且常會搭配於交叉驗證的方式進行超參數搜索。而分類器的機率切點則也用於相同方式，進行尋找出最佳機率切點，以達到最高的預測效果。

### 醫學檢測觀點

在醫學中常藉由人工智慧進行初步的醫療檢測，為了初步找尋所有罹患的人，通常醫師會採取「高敏感度」的切點(臨界值較低的)，寧可將部分正常人被誤判，也不想錯失罹患者的急救時間。所以初步是由人工智慧的簡易檢測，之後才會由醫師的精密檢查過濾被誤判的正常人。