

分類器的機率切點[◎]

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

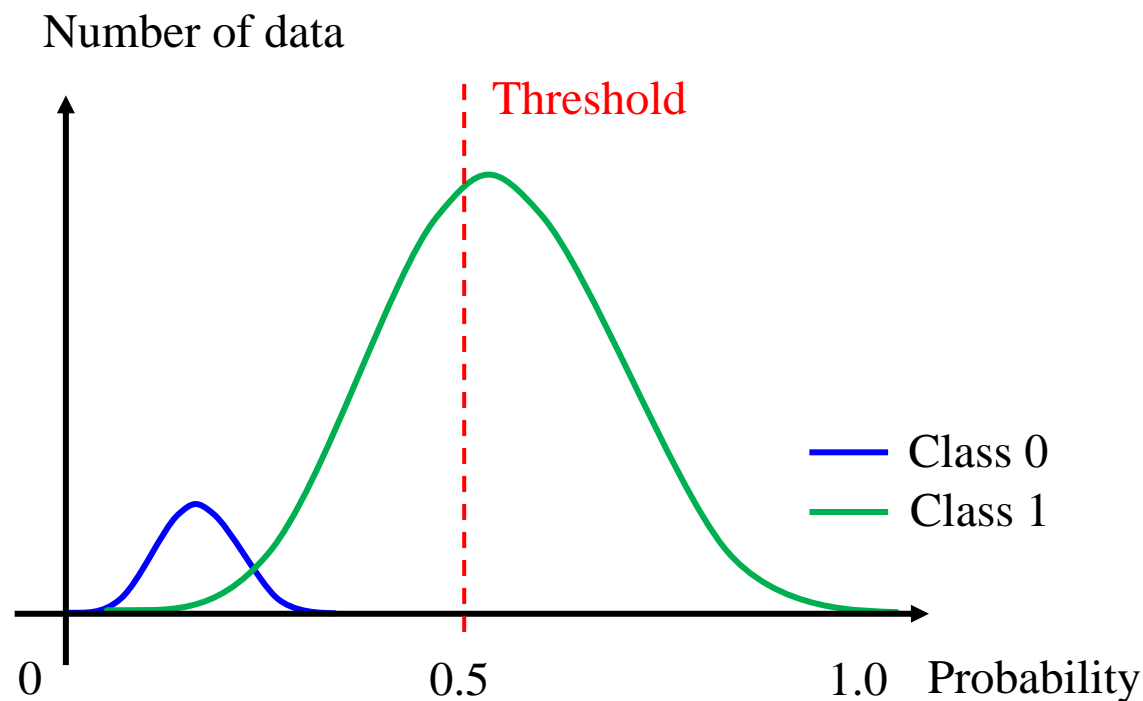
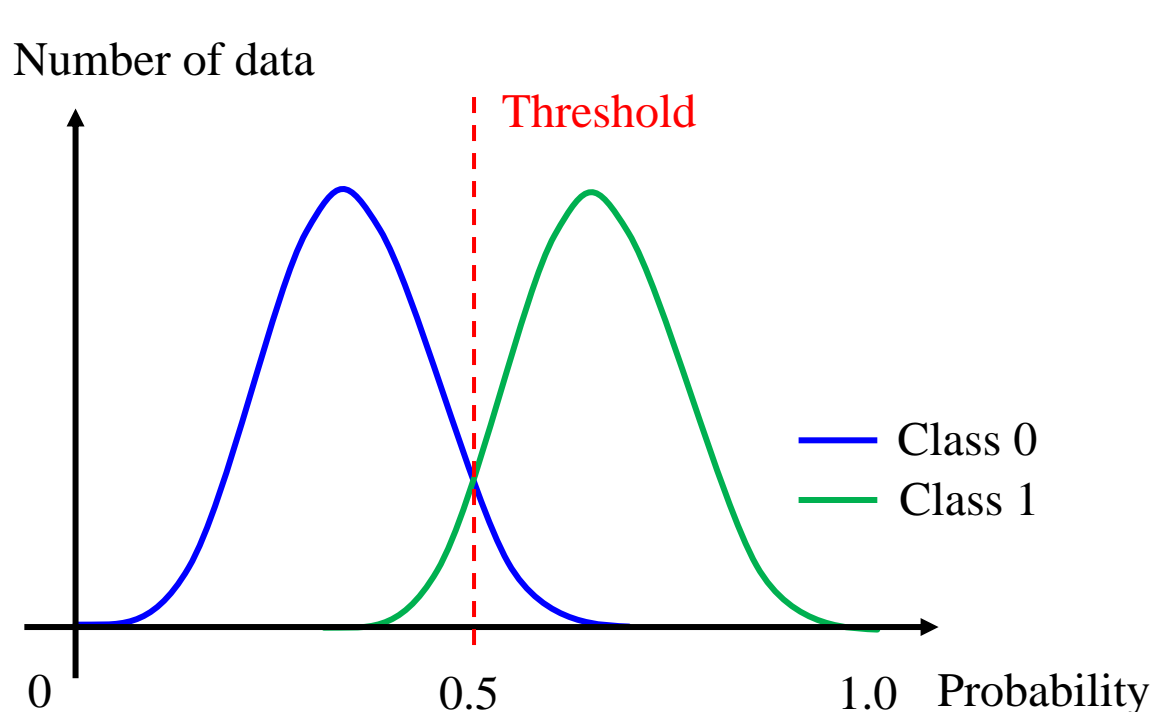
3

4

5

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



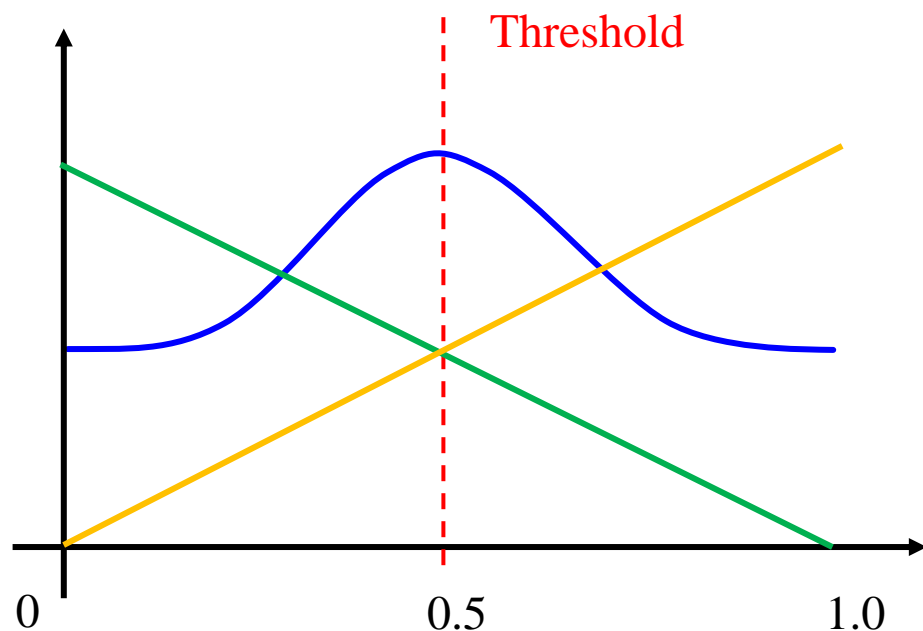
分類器的機率切點

02.指標的選擇

📄 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

