

2019 ECT 作業一

1. 用 Weka 軟體對 diabetes.arff 利用 Naïve Bayes 進行 Supervised learning, 選擇 “Use training set”, 設定 Attribute: class 為 Output, 在過程中對重要步驟截圖加以說明, 並回答以下問題:

(a) 解釋 Classifier Output, Test data 的錯誤率是多少? 有多少百分比的 Test dataset instances 被分類到 tested_negative class 但實際上屬於 tested_positive class? 請利用 Confusion matrix 解釋。(15%)

(b) 在 Output predictions 的結果中, 欄位 error 出現 “+” 代表意思為何? 請截圖並解釋之。(10%)

(c) 請利用 Visualize Classifier Errors, 找出預測錯誤的資料點 3 個, 並寫出各是第幾筆資料, 請截圖操作步驟並解釋。(15%)

(d) 請使用 Visualize Classifier Errors, 解釋產生的圖以及此圖與 Confusion matrix 之間的關係。(10%)

2. 用 python 對 diabetes.csv 進行 Supervised learning 中的 Naïve Bayes 分析, 並回答以下問題:

(a) 請問 diabetes 各屬性('preg', 'plas', 'pres', 'skin', 'insu', 'mass', 'pedi', 'age') 的平均值各為何? (5%)

(b) 在過程中對所有重要程式步驟進行截圖並加以說明, 越詳盡越好。(10%)

(c) 請利用 metrics.classification_report() 呈現出最後 precision. recall. f1-score 值, 並截圖加以說明。(10%)

(d) 請利用 metrics.confusion_matrix() 呈現出混淆矩陣, 並截圖加以說明。(10%)

(e) 請問當 'preg'=2, 'plas'=1, 'pres'=0, 'skin'=0, 'insu'=2, 'mass'=1, 'pedi'=2, 'age'=20 時, 最終的 output class 為何。(5%)

(f) 請比較 weka 和 python 分析之結果, 並加以說明。(10%)

- 繳交期限: 3/6(三) 中午 12:00
- 請轉檔為 PDF 格式, 檔名為:ECT_HW1_學號_版本.pdf,
並同時附上 python 的 ipynb.檔, 命名格式同上。
- 上傳至 LMS 作業區, 遲交一天扣該次作業成績 5%
- 補交請上傳至 HW1 補交區,僅開放到 3/13(三) 中午 12:00