

# 2020 ECT 作業六

1. 使用 weka 對 Social\_Network\_Ads.csv 進行 naive bayes 分析，選擇 percentage split 66% ，過程中對所有重要步驟進行截圖並加以說明，越詳盡越好：

(a) 使用 cost-sensitive-learning ，將 cost matrix 設定如下圖，列出 total cost 及 average cost ，截圖並詳細說明該數字是如何計算出來的 (20%)

0.0	5.0
3.0	0.0

(b) 對購買者 (purchase = 1) 進行 cost/benefit analysis ，cost matrix 一樣設定如下圖，說明最佳的 sample size rate 是多少？截圖並詳細說明 (20%)

Predicted (a)		Predicted (b)	
0.0	5.0	Actual (a)	
3.0	0.0		
		Actual (b)	

(c) 承上題，在最佳的 sample size rate 情況下，混淆矩陣長怎樣？cost 為多少？截圖並詳細說明該數字是如何計算出來的 (10%)

2. 使用 python 對 Social\_Network\_Ads.csv 進行 naive bayes 分析，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：

(d) 設  $\text{test\_size} = 0.33$  ,  $\text{random\_state} = 1$  , 進行 naive bayes 分析後, 列出準確率及 TP Rate/FP Rate (10%)

(e) 繪出 ROC Curve 並計算出 AUC (20%)

(f) 繪出 lift chart (又稱 Cumulative Gain Chart) (X軸: sample size rate; Y軸: TP rate) (10%)

(g) 繪出 lift curve (X軸: sample size rate; Y軸: Lift) (10%)

- 繳交期限: 5/20(三) 中午 12:00
- 請轉檔為 PDF 格式, 檔名為: ECT\_HW6\_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 HW6 補交區, 僅開放到 5/27(三) 中午 12:00