

1.請比較你實作的 generative model 、logistic regression 的準確率，何者較佳？

答：

	public	private
logistic regression	0.85515	0.85124
generative model	0.84471	0.84313

logistic regression 的準確率在兩個分數都較高

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

best model 上使用 scikit-learn，選用 AdaBoostClassifier，n-estimator 設為 1200，選用 SAMME.R 演算法，accuracy 在 public 達到 0.87285，private 則是 0.87139

3.請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響。

答：

	public	private
logistic regression (fn)	0.85515	0.85124
logistic regression	0.80614	0.80002
generative model ()	0.84471	0.84313
generative model	0.84557	0.84277

(fn)表示有做 feature normalization

看的出來在 logistic regression 上差別較大，但在 generative model 幾乎沒影響

4. 請實作 logistic regression 的正規化(regularization)，並討論其對於你的模型準確率的影響。

答：

實作上把 regularization 設為 0.01，在此比較 0 0.1 以及 0.01 的 performance

regularization	public	private
0	0.85515	0.85124
0.1	0.85515	0.85112
0.01	0.85515	0.85124

以上述參數來看，regularization 影響並不大

5.請討論你認為哪個 attribute 對結果影響最大？

fnlwgt 對結果影響似乎不大，去除後結果略有上升，此外在國家的部份我有人為分成 4 個 class，降維之後準確率也略有上升

以 normalize 後的權重來看，Seperated(分居)有-0.712，應該是影響最大的 attribute