

學號：B04901015 系級：電機三 姓名：傅子興

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

答：由我做出的兩個模型進行比較，以logistic regression的準確率較高

以下為兩者於kaggle 上測試的成績：

	generative model	logistic regression
accuracy	0.82432	0.85491

2.請說明你實作的best model，其訓練方式和準確率為何？

答：我的用SciKit 中的AdaBoostClassifier (需要搭配DecisionTreeClassifier使用)

我的模型架構是：Algorithm設定SAMME.R, n_estimator設定1000

根據我的測試，隨著weak learner數目越多，結果越精確，符合理論

根據kaggle上顯示的成績，其準確率為87.272%

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：有做normalization的話正確率會上升，以我自己的碼做測試，得出下表的正確率比較
(註：iteration設定10000，二者皆有regularization, $\lambda=0.01$)

	without normalization	normalization
accuracy	0.79903	0.85253

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：綜合測試結果，最佳 λ 發生於 $\lambda = 0$ or 0.001

regularization(λ)	accuracy
0	0.85271
0.1	0.85253
0.01	0.85259
0.001	0.85271
0.0001	0.85259

從我測試的結果來看，普遍上 λ 小一點正確度較高，但在到0.001時又上升，關係不明確

5.請討論你認為哪個attribute對結果影響最大？

答：

雖然並沒有實際將參數拿掉進行測試，但是經由對training data 跟testing data的比對跟分析關聯性之後，我個人覺的國家影響非常大；一件令人意外的是學歷好像影響並沒有想像中大，關聯不強