

1. 請說明你實作的 generative model，其訓練方式和準確率為何？

採用投影片中的示範的高斯分布，並共用 Sigma ($\text{Sigma} = P_C0 * \text{sigma}_0 + P_C1 * \text{sigma}_1$)。Kaggle public 準確率約為 83.7

2. 請說明你實作的 discriminative model，其訓練方式和準確率為何？

採用 Logistic Regression，用助教整理好的 X_train 為納入所有 106 個特徵，使用 SGD 配上 ADAM 來訓練資料。Iteration 設定為 75。最後 kaggle public 分數準確率約為 85.399%。

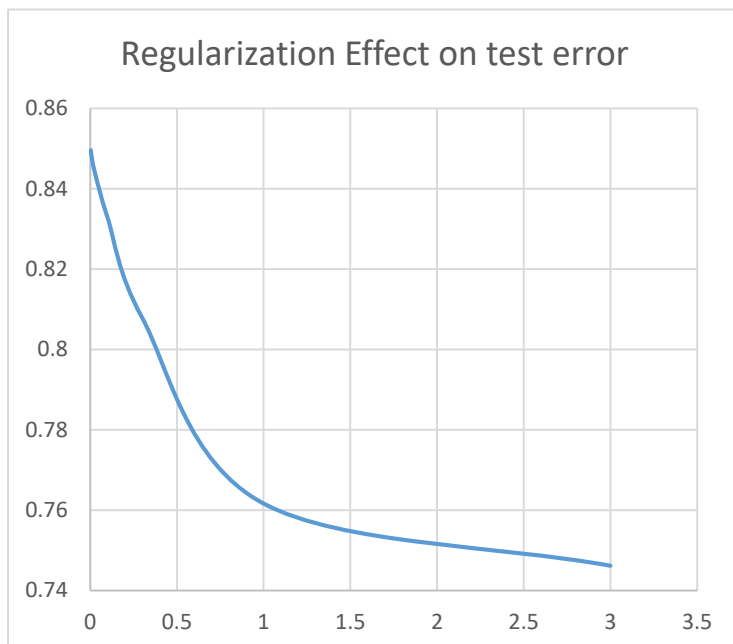
3. 請實作輸入特徵標準化(feature normalization)，並討論其對於你的模型準確率的影響：

標準化後的資料會大幅減少訓練所花費的 iteration。未標準化的資料大概需要 20 倍的 iteration 才能達到 85%。

Without feature normalization : 0.742239184187

With feature normalization : 0.850004572499

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。



Lamda	Accuracy
3	0.746138
1	0.761617
0.3	0.807837
0.1	0.832775
0.03	0.843432
0.01	0.847456
0.003	0.849605
0.001	0.849728
0.0003	0.850343
0.0001	0.850834
0.00003	0.849974
0.00001	0.850189

5. 請討論你認為哪個 attribute 對結果影響最大:

以我訓練出來的 weight 來看，總體影響最多的因素為 capital gain。越高的 capital gain 則代表越有機會年薪超過 50k。