

學號：R06521608 系級： 土木碩一 姓名：陳德元

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

答：很意外的 generative model 好像較佳。

	Generative model	Logistic regression
Valid acc	0.838145	0.789285
Public	0.84533	0.78832

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：用 sklearn 實作 logistic regression，並且用 sfs 演算法，刪掉其中 7 個變數，分別是 'fnlwgt'、'Prof-school'、'Adm-clerical'、'Greece'、'Trinidad&Tobago'、'United-States'、'Vietnam'，其 public score = 0.85466

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：皆提升準確率。

	Generative model	Logistic regression
Valid acc	0.845516	0.823545
Public	0.84484	0.82272

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：好像沒有特別變高，反而變低。

Validation acc = 0.804880

Public = 0.79754

5.請討論你認為哪個 attribute 對結果影響最大？

答：capital_loss，因為在用 sfs 找 attribute 的時候，只要 capital loss 一個 attribute 就可以有 80%的正確率

