學號：R06922152 系級： 資工碩一 姓名：袁晟峻

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

答：

Generative model：kaggle public 0.84533

Logistic model：kaggle public 0.84680

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

用keras的sequential model 實作三層架構的NN，第一層106個neuron，activation 用 relu，第二層64個neuron，activation 用 relu，第三層1個neuron，activation 用 sigmoid，optimizer用adagrad，第一層到第二層中間Dropout(0.5)，第二層到第三層時也Dropout(0.5)，用這樣的model去做訓練，在kaggle上public分數為0.85823。

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：

在generative model 的部分，特徵標準化對準確率的影響感覺不大，我測得的training accuracy都在0.84多

在logistic model的部分，若是不做特徵標準化，在算sigmoid使用numpy.exp的時候會overflow，導致model訓練不起來，做完標準化之後訓練完的model可達到0.84多的準確率

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：

5.請討論你認為哪個attribute對結果影響最大？

Race 和 Age