學號：r05942114　系級： 電信一　姓名：方敬勻

1.請說明你實作的 generative model，其訓練方式和準確率爲何？
答：
使用 106 個由助教取出的 train data 中的 feature 在訓練時沒有做過 Normalization，直接使用 generative 中的高斯 model 做訓練，做出來的準確度在 Kaggle 中爲 0.83 左右。

2.請說明你實作的 discriminative model，其訓練方式和準確率爲何？
答：
在 discriminative model 當中使用的 Data 經過 Normalization，在 training data 時使用 adagrad 以及 regulization，使用 adagrad 時讓 learning rate 的調整變爲較爲簡單較不容易出現 overflow 的情形，在 di
scriminative model 中我取出每 5000 筆 data 爲一 data set 跑完 model 後在與全部 3 萬筆 data 做錯誤率分析出來之後約爲 85%的正確率與 Kaggle 的結果差不多。

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：
在使用 feature normalization 之前在取 log 時容易發生問題，在做 feature normalization 時也會需要注意一些除數爲 0 的狀況，做 Normalization 前後的 Error rate 相差非常的顯著在做 Normalization 之前錯誤率大概爲 75%，在 Normalization 之後提升到 85%左右。

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：
在使用 Regularization 時我的 lambda 調在 0.0001 在使用 Regularization 時在 Train data 中的 Loss function 會比較高一點跟未使用 Regularization 相差約爲 2%左右，不過在與未使用 regularization 的 data 同樣輸入全部 Train data 做測試時錯誤率沒有明顯的差別，錯誤預測的資料在 3 萬筆 Data 中相差使有十位數而已。

5.請討論你認爲哪個 attribute 對結果影響最大？
這邊我使用各別 Feature 的 Weighting 來判斷相關程度選出的參數爲資本收益,不曾工作,以及國籍從人的角度思考這三個參數與薪水的確會有非常大的相關性。