學號：r06946004 系級： 資科學程碩一 姓名：蔡尚錡

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

答：

經由實驗，我發現我的logistic regression的準確率普遍能比generative model高出比較多，generative model 大概落在public = 0.847 private = 0.843左右 ，而logistic regression 經由一些超參數的調整可以達到public = 0.856 private = 0.850左右，因此我的logistic是比較佳的。

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

我的best model是用tensorflow從底層實作classification task的 Neural Netwok，兩層的hidden layer ，每一層大小是1024，output layer 大小是2，每一層裡面的一個neuron 其實就是一個logistic regression，經由大量的neuron疊加在一起讓參數量大幅增加之後來訓練，利用backpropagation 更新參數，搭配AdagradOptimizer，activation function 我用內建的relu6，learning rate = 0.01，batch_size = 32，這些參數都是慢慢調過以後的結果，一開始hidden layer設256或512都太小結果不盡理想，直到把hidden layer開大之後才能確實學到好的feature，batch_size設太大也會導致結果下降，relu 也比 sigmoid結果較佳，最後訓練出來超過strong baseline 到達public=0.857 private=0.851左右，或許是因為我的 feature 沒有特別挑過，也或許是超參數沒有挑到最好，才導致沒有特別高的準確率，但用 Neural Network 方法卻比一般的logistic方法在privacy testing_set表現的還要穩定。

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：

我用training set 裡面隨機取出10%當作validation set ，feature 一樣是直接使用助教預先抽好的，訓練次數為1500 次，初始學習率為0.01，並且使用adagrad

Normalization?	Training (accuracy/loss)	Valid (accuracy/loss)
Yes	0.85035/0.36693	0.85433/0.36430
No	0.80843/0.47212	0.81450/0.46969

可以看到在相同的epoch和learning rate之下，feature normalization 能大幅提升準確度和效能。另外，若沒有使用 feature normalization ，即使用了adagrad，初始學習率也必須好好選擇，否則容易在訓練過程中發散。

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：

origin為助教預先抽取好的 feature， square 為age，fnlwgt，capital gain ，capital loss，和hours per week，這五個欄位的平方項。

Regularization constant	Feature	Training (accuracy/loss)	Valid (accuracy/loss)
0.0	origin	0.85390/0.31687	0.85176/0.31259
0.001	origin	0.85345/0.32714	0.85205/0.32387
0.0001	origin	0.85371/0.31809	0.85176/0.31471
0.0	origin+square	0.85678/0.30789	0.85699/0.30716
0.001	origin+square	0.85548/0.32699	0.85644/0.32449
0.0001	origin+square	0.85639/0.31042	0.85623/0.31071

從實驗數據觀察看到，加了regularization後，對模型準確率的影響並不大，大概都增加準確率0.004左右，也許是feature的選得不太好，使得模型並沒有太過複雜，因而不會與訓練資料過度擬合，產生overfitting。

5.請討論你認為哪個attribute對結果影響最大？

單單討論一次項模型進行比較，試著把各項attribute拔掉後，觀察其結果，可以發現移除capital gain後的結果最差。原先的0.85左右的準確率會因此掉到0.83左右，其他attribute移掉大概還維持0.85或略低0.002左右。並且在我將各個attribute進行各種不同方式的轉換以觀察其實驗結果時，加上 log(capital gain) 後，收斂速度快很多，在 public set 上也能取得較好的成績，因此我認為是capital gain影響最大。