

學號：B04901066 系級：電機三 姓名：洪國曉

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

答：

model	Private Score	Public Score
generative	0.84191	0.84606
logistic	0.80862	0.79975

generative model的準確率較佳(高近5%)。(皆有feature normalization)

2.請說明你實作的best model，其訓練方式和準確率為何？

答：

使用keras(如右圖)

將 X_train normalize 至 0~1

準確率：

Private Score：0.84498

Public Score：0.85528

```
Using TensorFlow backend.
Layer (type)                 Output Shape          Param #
=====
dense_1 (Dense)              (None, 1024)          109568
dropout_1 (Dropout)          (None, 1024)          0
dense_2 (Dense)              (None, 1024)          1049600
dropout_2 (Dropout)          (None, 1024)          0
dense_3 (Dense)              (None, 1024)          1049600
dropout_3 (Dropout)          (None, 1024)          0
dense_4 (Dense)              (None, 1)             1025
=====
Total params: 2,209,793
Trainable params: 2,209,793
Non-trainable params: 0
```

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：

model	Private Score	Public Score
generative有normalization	0.84191	0.84606
generative無normalization	0.84289	0.84484
logistic有normalization	0.80862	0.79975
logistic無normalization	0.78774	0.79176

generative幾乎無影響，因其直接用公式計算，所以不會受到參數scale影響。

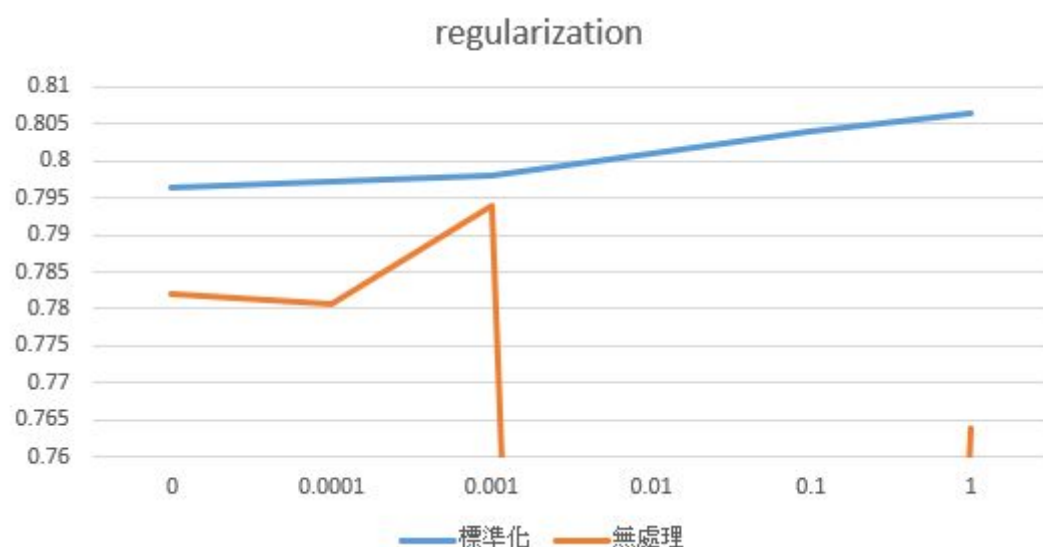
反之logistic就有十分明顯的影響，甚至在訓練時，發生準確度在0.76和0.24之間跳的特殊狀況。

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：

由下列表格和作圖可知，標準化後適度的正規化可提升模型準確率，無處理基本上就是壞掉了。

lambda	標準化	無處理
0	0.796385	0.781946
0.0001	0.797305	0.780598
0.001	0.797981	0.794053
0.01	0.801111	0.236225
0.1	0.803881	0.236225
1	0.806525	0.763771



5.請討論你認為哪個attribute對結果影響最大？

從標準化後的logistic regression model的w參數，可發現助教提供之X_train，第四欄之參數影響最大，即capital gain的值，如果只用capital gain是否>5000來判斷，就有 Private Score：0.79977；Public Score：0.80405這樣很不錯的表現，幾乎跟logistic regression model表現一樣，因此我認為capital gain對結果影響最大。