

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

答：

	Validation	Public	Private
Generative model	0.842111	0.84520	0.84240
Logistic regression(5000次)	0.853321	0.85442	0.85112
Logistic regression(50000次)	0.853321	0.85454	0.85112

就結果來看，logistic regression的預測準確率要比generative model好一點，不過regression的參數有很多可變動的部分，所以並不能保證logistic regression一定比generative model要來的好

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訓練NN，準確率為public：0.86159

model部分，實作三層hidden layers(512, 512, 256)，前兩層用relu，第三層用sigmoid，在第一層的部分加入L1的regularize( $\lambda = 0.001$ )，optimizer使用Adam(lr = 0.001)

feature部分，我有自己挑選過feature，並把部分feature加2次項跟3次項

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：

以下model 都有做feature normalization

	Validation	Public	Private
Generative model	0.842296	0.84545	0.84240
Logistic regression(5000次)	0.853321	0.85442	0.85112
Logistic regression(50000次)	0.853321	0.85454	0.85112

在generative model的部分，標準化後的accuracy只比原本的高一點點，對於generative model來說，不管有沒有標準化都可以做得起來，在logistic regression方面的話，不做標準化會train不起來，因此我的logistic 數據都是有經過標準化的data做出來的

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：

	Local Accu	Public	Private
$\lambda = 0.1$	0.853321	0.85075	0.85417
$\lambda = 0.01$	0.853321	0.85112	0.85454
$\lambda = 0.001$	0.853321	0.85112	0.85454
$\lambda = 0.0001$	0.853321	0.85112	0.85454

對於不同的 $\lambda$ 而言, accuracy並沒有差多少, 可能是因為training次數還不夠多, 還沒到達over fitting

5.請討論你認為哪個attribute對結果影響最大？

All : 0.90811

-workclass : 0.89594

-education : 0.881913

-marital\_status : 0.90611

-occupation : 0.88492

- relationship : 0.900924

-race : 0.902889

-native\_country: 0.904855

-age : 0.893000

-fnlwgt : 0.90083

-sex : 0.90522

-gain : 0.89536

-loss : 0.90424

-hours : 0.90046

以上是我對拿掉不同attribute所做的validation accuracy, 發現拿掉education跟occupation後所得accuracy下降最多, 因此我認為education跟occupation是對結果影響最大的attribute