

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

答：

用 private + public 的 accuracy 來看:

Generative: 0.8438056630428107

Logistic: 0.8218

Generative 有做 normalization，Logistic 沒有

Generative 結果較好

2.請說明你實作的 best model，其訓練方式和準確率為何？

答：

Best model 使用 keras 實作 NN，共設兩層 hidden layer，第一層設 60 個 neural，activation function 為 sigmoid，第二層設 40 個 neural，activation function 也是 sigmoid，最後一層 output 的 activation function 也使用 sigmoid，並對 training data 作 normalization，第二層的 hidden layer 作 regularization，optimizer 使用 adam，batch\_size 為 9000，切 0.33 的 data 當 validation set，跑 600 個 Epoch，最後 public + private 的 accuracy = 0.85633。

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：

用 private + public 的 accuracy 來看:

沒做 feature normalization 的 accuracy = 0.85375，有做 feature normalization 的 accuracy = 0.85633，有做的比沒做的略高，可能是因為 normalization 可以加快 training 的速度，但在這個 model 下 training 到最後可能已經到 model 接近的最佳解，所以兩者的 accuracy 定沒有差太大。

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：

兩者都沒做 normalization，因為做了對於 model 沒有什麼的改善

用 private + public 的 accuracy 來看:

No regularization: 0.8195

Regularization: 0.8218

對 model 做 regularization 後，accuracy 合理有提升，但只有提升少量。

5.請討論你認為哪個 attribute 對結果影響最大？

fnlwgt，算出所有 feature 與 Y 的相關係數，再由相關係數絕對值最小的開始嘗試，發現將 fnlwgt 加入後 accuracy 下降到 0.7811559，相較於其他的 feature 下降最多。