

## HW2\_Report

學號：B05901005 系級：電機三 姓名：賴沂謙

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

Logistic regression 能調整的參數較多，準確率也較佳；Generative model 的準確率則有一定的落差。

2. 請說明你實作的 best model，其訓練方式和準確率為何？

Feature：使用 X\_train 裡面抽好的 feature，但是拿掉 fnlwgt 以及所有的 Nation，再加上原始 train.csv 裡面的 education\_num 還有把 age 分成 classes，最後將所有 continuous 的項加上平方項及立方項並且做 normalization。

Train：使用 regularization  $\lambda=0.01$ ，epochs=3000，learning rate=1，Adagrad 訓練。

Accuracy：public = 0.85847, private = 0.85861

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

		Public Accuracy	Private Accuracy	Average
Gene-rative	Raw data	0.82764	0.82115	0.824395
	$x' = \frac{x - \text{mean}}{\text{std}}$	0.81719	0.80628	0.811735
Regre-ssion	Raw data	0.78685	0.77975	0.7833
	$x' = \frac{x - \text{mean}}{\text{std}}$	0.85847	0.85861	0.85854
	$x' = \frac{x - \min}{\max - \min}$	0.85700	0.85702	0.85701

對於 generative model 而言，因為都是算數據本身的統計值，所以做了 normalization 效果並沒有比較好。然後對 logistic regression 而言，normalizaiton 就差很多了，因為這次每個 feature 的值都相差很多，有些數字都是幾千萬，而 discrete 的 feature 都是 0 跟 1，沒有經過 normalization 就直接 train 就會有問題。

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。

	Public Accuracy	Private Accuracy	Average Accuracy
$\lambda=0$	0.85835	0.85861	0.85848
$\lambda=0.1$	0.85835	0.85886	0.858605
$\lambda=0.01$	0.85847	0.85861	0.85854
$\lambda=0.001$	0.85847	0.85861	0.85854

使用 regularization 後，正確率會上升一點點，應該是因為這次 feature 有使用到平方及立方項，其中  $\lambda=0.1$  的正確率稍微高一點。

5. 請討論你認為哪個 attribute 對結果影響最大？

ATTRIBUTES	W
PRESCHOOL	-11.771308
WITHOUT-PAY	-6.4643743
NEVER-WORKED	-5.34925
PRIV-HOUSE-SERV	-3.9040729
AGE_0-19	-2.8500046
AGE_90-	2.78964501

根據 train 完的 model 來看，Education 裡面的 Preschool 負影響最大，再來就是 Work Type 裡的 Without-Pay、Never-Worked 跟 Priv-House-Serv 也同樣是對結果造成負影響，還有年齡的部分也影響了蠻大一部份。