

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

答：

方法	準確率 (public / private)
generative	0.84252 / 0.84496
logistic	0.84964 / 0.85233

我有在 logistic 中加入 adagrad

但其實好像沒什麼太大的差別

以上表的結果來看，不論是 public 或是 private，都是 logistic 的方法比較好一些

但是單以我作出的結果來看 logistic 也沒有比 generative 高出太多

我覺得有可能是 logistic 的參數沒調好

2.請說明你實作的 **best model**，其訓練方式和準確率為何？

答：

方法	準確率 (public / private)
generative	0.84252 / 0.84496
logistic	0.84964 / 0.85233
DNN	0.84633 / 0.85491

我的 best model 是用了 Keras 去建構 neural network

只是用了五層很簡單的 fully connected 層

每層 20 個 unit

performance 相較於單純的 logistic regression 有一些成長

本來單純只用 logistic 的方法是沒有過 base line

後來用了 deep learning 之後就勉強壓線過 strong base line

但是後來看了 private 之後發現是 logistic 的 private 比較高

所以感覺我實作出的 DNN 跟 logistic 好像差不多

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：

方法	準確率 (public / private)
Logistic with normalization	0.77982 / 0.78532
Logistic without normalization	0.84964 / 0.85233
generative with normalization	0.84252 / 0.84496
generative without normalization	0.84264 / 0.84557

從上表可以看出 logistic 沒有 normalize 的準確度十分的差
但是 generative 卻不然
去掉 normalization 之後 generative 的準確度卻上升了
所以 normalization 應該不適用於 generative

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：

方法	準確率 (public / private)
$\lambda = 0.1$	0.83319 / 0.83562
$\lambda = 0.01$	0.83429 / 0.83230
$\lambda = 0.001$	0.83024 / 0.83255

差距不太大
不過其實還是可以觀察到 λ 下降時
準確度也會跟著下降

5.請討論你認為哪個 attribute 對結果影響最大？

我有測試過拿掉一些 attribute
感覺好像是用上全部的 attribute 會最好
如果把一些拿掉的話準確度通常都會下降一點
但也沒有下降太多
印象中把 capital gain 那個部分拿掉影響會最大
Validation 會下降到 0.82 左右
所以我覺得 capital gain 的影響還蠻大的