

**1. 請說明你實作的 generative model，其訓練方式和準確率為何？**

答：我的 Generative Model 是使用 Gaussian Distribution，直接帶入教授投影片中  
的公式，因為在實作過程中有發現 class0 或是 class1 其一的 covariance matrix 會  
是 singularity matrix，所以共用 Covariance matrix 來避開這種狀況，並期望能向  
教授投影片中的結果一樣提高準確度，由於  $p(x|C_0)$  和  $p(x|C_1)$  在計算的公式中有  
exponential，導致機率為 0，所以我並未計算  $\frac{p(x|C_0)}{p(x|C_0)+p(x|C_1)}$ ，而是直接比較  $p(x|C_0)$   
和  $p(x|C_1)$  的大小，若  $p(x|C_1) > p(x|C_0)$ ，則 output 為 1。實作結果在 training set  
上 accuracy  $\approx 84.1927\%$ ，在 public leader board 上的 accuracy  $\approx 84.103\%$ 。

**2. 請說明你實作的 discriminative model，其訓練方式和準確率為何？**

答：我使用助教抽取好的 feature，將所有 feature 都做 feature scaling，並使用  
二次是 (即  $\sigma(w_2x^2 + w_1x + b)$ )，搭配 stochastic gradient descent，mini-batch  
size=50，initial learning rate=1，每跑 200 個 iterations 就把參數存下來，再手  
動調整 initial learning rate，使得再 train 的時候 accuracy 能有一定效率的上  
升，到最後 accuracy 感覺不太上升之後，就停下载式，改程式碼不再存新的參數，  
讓程式再跑一次，人工看 accuracy 最高的點停下來，在把程式中之 iterations 數改  
成此數值，再次重跑一次程式，即得到最佳的結果。最後我的 accuracy 如下表，

	Training set	Public leader board	Private leader board
Accuracy	85.84810%	85.737%	85.788%

**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：我比較  $\sigma(w_2x^2 + w_1x + b)$  和  $\sigma(w_1x + b)$  有無 normalization 的結果，並使用教授  
投影片中 feature scaling 的方式 ( $z_i = \frac{x_i - \mu_x}{\sigma_x}$ ) 對 input  $X_{traincc}$  和  $X_{test}$  做  
feature normalization。

Training/Public leader board	$\sigma(w_2x^2 + w_1x + b)$	$\sigma(w_1x + b)$
有 normalization	85.84810% / 85.737%	85.3536% / 85.356%
無 normalization	79.5525% / 79.730%	85.34136% / 85.356%

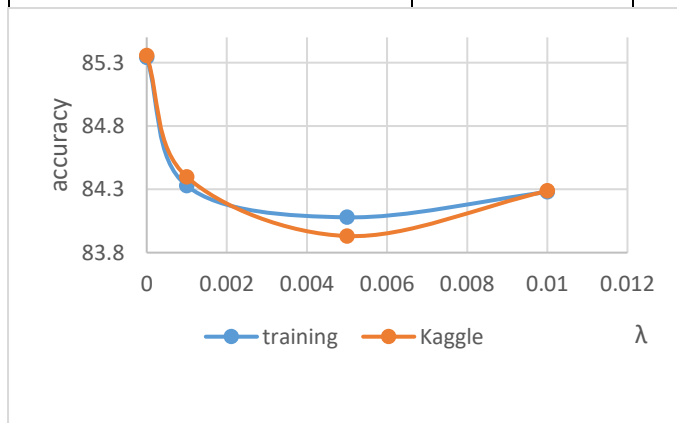
加入 feature normalization 有兩個好處：

- (a) 在 trining 時，Loss 下降比較快也比較穩定
- (b) 由表中可以得知，當使用的 function 是  $\sigma(w_2x^2 + w_1x + b)$ ，加入 normalization  
不論 training set 或是 public leader board 上的準確率都有顯著的上升，若  
function 是  $\sigma(w_1x + b)$ ，加入 normalization 之後，只有 training set 的準確  
率有上升。整體而言，加入 feature normalization 可使模型的預測準確率上  
升。

**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：我比較一次式  $\sigma(w_1x + b)$  沒有做 normalization 的條件下，不同  $\lambda$  下的結果。

$\lambda$	0	0.001	0.005	0.01
Training (%)	85.34136	84.3278	84.079	84.2787
Public leader board (%)	85.356	84.398	83.931	84.287



相較於作業一，隨著在 training set 上預測準確率越高，可能越容易 overfitting，而造成在 public leader board 的準確率下降，加入 regularization 是為了避免 overfitting，而且會使得 model 在 training set 上的表現結果差一點，以求在 public leader board 上準確率能夠提升，但在作業二中 training set 的準確率與在 public leader board 上的準確率大致上是呈現正相關性，加入 regularization 後會讓 training set 上準確率表現大幅下降，即使在 public leader board 的準確率有機會上升，但上升幅度都不大，因此整體而言加入 regularization 並無法提升準確率，反而會讓準確率下降！

##### 5. 請討論你認為哪個 attribute 對結果影響最大？

答：我比較使用助教幫我們的 feature，只是用單一種 attribute 在 training set 和 public leader board 的結果，如下表：

	age	fnlwgt	sex	capital - gain	capital - loss	hours-per-week
training	75.9190%	75.9190%	75.9190%	80.0928%	77.2489%	75.9190%
public leader board	75.8110%	76.0320%	76.0320%	80.0610%	77.2360%	76.0320%

work-class	education	marital-status	occupation	relationship	race	native-country
76.3122%	77.9583%	75.9190%	75.9190%	75.9190%	75.9190%	75.9190%
76.5110%	77.4820%	76.0320%	76.0320%	76.0320%	76.0320%	76.0320%

由上表可知對預測結果影響最大的應會是：capital - gain > education > capital - loss > work - class，剩下的 attribute 就有差不多的影響，但若只用 capital - gain, education, capital - loss, work - class 做預測，結果仍然沒有用全部 attribute 來的好。