

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

答：我是利用上課老師推導 Gaussian distribution 的公式，取所有助教處理過後的 features，以相同的 Σ ，不同的 μ 直接求出 ω 和 b 。正確率為 0.84152 (kaggle 上的成績)。

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

答：我將 features 先做 normalization，取國家以外的一次式，再加上 age、fnlwgt、sex、capital_gain、capital_loss、hours_per_week 的五次式，以 sigmoid 作為 activation function 去訓練我的 model。正確率為 0.85995 (kaggle 上的成績)。

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：若先對 features 做 normalization 再取高次能達到較高的正確率，我的推測是做完 normalization 後的高次 features，在算 exponential 的時候較不容易產生 overflow 等情形，因此在訓練的過程有較佳的 performance。

normalization	accuracy
不做 normalization	0.7591904425539756
先做 normalization 再取高次	0.8589109671078898
先取高次再做 normalization	0.7591904425539756

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：當 λ 值小時，到正確率有些許幫助，隨著 λ 越來越大，正確率則越來越低。(以 training set 的正確率)

λ	accuracy
0	0.8577132151960934
0.1	0.8577746383710574
0.5	0.8577439267835755
1	0.8576825036086115
10	0.8576517920211296
100	0.850925954362581
1000	0.7723042904087712

5.請討論你認為哪個 **attribute** 對結果影響最大？

答：我覺得 age、sex、capital_gain、capital_loss、hours_per_week 這五個 features 對結果的影響較大。若將這幾項抽離的話，正確率會大幅下降，而又以 capital_gain 的影響最大。