

學號：R05922096 系級：資工碩一 姓名：李哲安

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

答： step1. 取所有 feature 並轉為 3 次方 data  $[x_1, x_1^2, x_1^3, x_2, \dots]$

step2. 做 feature normalization

step3. 計算出 mean1, mean2 和 covariance matrix

step4. 根據投影片公式轉為 w 跟 b

$$z = \frac{(\mu^1 - \mu^2)^T \Sigma^{-1} x}{w^T} - \frac{\frac{1}{2} (\mu^1)^T \Sigma^{-1} \mu^1 + \frac{1}{2} (\mu^2)^T \Sigma^{-1} \mu^2 + \ln \frac{N_1}{N_2}}{b}$$

kaggle 上的分數為 0.84177

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

答： step1. 取所有 feature 並轉為 3 次方 data  $[x_1, x_1^2, x_1^3, x_2, \dots]$

step2. 做 feature normalization

step3. Train by AdaDelta

kaggle 上的分數為 0.85577

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： 將所有 feature 減去 mean 之後除以(max - min)

若(max - min) == 0 則除以 1

kaggle 上的分數由 0.83256 上升到 0.85577

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： 實做 regularization 後，準確率反而下降了(0.85577 => 0.85049)。

由於原本在 training set 上的準確率跟 testing set 上的準確率並沒有落差太大，推測原本並沒有 overfitting 的情況，所以加上 regularization 之後反而限制了 model 對 training data 的 fitting 導致準確率下降。

5.請討論你認為哪個 attribute 對結果影響最大？

答： Prof-school 的影響最大。

因為若只取一個 feature 做 generative model，取 Prof-school 時會得到最高的準確率，再來依序是Masters, capital\_loss。