

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

---

答：

實做上課提到的 Gaussian Model，

$$P(C_1|x) = \sigma(w^T x + b); P(C_0|x) = 1 - P(C_1|x)$$

其中

$$W^T = (\mu^1 - \mu^0)^T \Sigma^{-1}; b = -\frac{1}{2}(\mu^1)^T \Sigma^{-1} \mu^1 + \frac{1}{2}(\mu^0)^T \Sigma^{-1} \mu^0 + \ln \frac{N_1}{N_0}$$

訓練方式為從訓練資料直接個別計算兩類（1 與 0）的平均值（ $\mu_1, \mu_2$ ），以及整體資料的共變異數（ $\Sigma$ ）。

準確度為 0.8412（使用  $\frac{2}{10}$  訓練資料作為驗證資料）。

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

---

答：

實做 Logistic Regression，並用 Adagrad Gradient Descent 與 mini batch 最大化模型的可能性。

準確度為 0.85565（Kaggle）。

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：

如果沒有標準化則梯度下降將變得十分困難，準確度只有 0.7635。有標準化後，準確度則可以到 0.85565。

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：

$\alpha$  為正規化項的係數。

$\alpha$	0	0.1	0.01
Accuracy	0.85334	0.85365	0.856470

因為本來過適的狀況就不太明顯，所以正規化對於準確度沒有太大的影響。

##### 5.請討論你認為哪個attribute對結果影響最大？

---

Capital Gain。經由測試每次只用單一個特徵訓練與預測，發現只用 Capital Gain 即可達到 0.8113 的準確度，至於其他特徵如果單一使用，則準確度都只有大約 0.74 ~ 0.77。