# ML 2019 Fall Hw2 Report

學號：b05705042  系級：資管四  姓名：皇甫立翔

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

   Logistic model 較佳。
   Logistic model：public score = 0.85577，private score = 0.85063。
   Generative model：public score = 0.83341，private score = 0.83773。

2. 請實作特徵標準化 (feature normalization) 並討論其對於你的模型準確率的影響。

   Logistic model 沒有做標準化：public score = 0.76254，private score = 0.76048
   Logistic model 有標準化：public score = 0.85577，private score = 0.85063
   Generative model 沒有做標準化：public score = 0.83341，private score = 0.83773
   Generative model 有標準化：public score = 0.76769，private score = 0.76538

   從結果可以觀察到說，logistic model 有做標準化，對於精確率來說提升幅度非常大。
   但是對於 generative model 來說，做了標準化反而會使精確率下降。

   我認為這是訓練方式的不同導致的，因為 generative model 有固定的公式，所以做了標準化反而會使結果變得較不精確。但是 logistic model 是透過不斷的訓練，所以標準化對於訓練就會非常有幫助。

3. 請說明你實作的best model，其訓練方式和準確率為何？

   我使用的是 Gradient Boosting Classifier，搭配 sklearn 裡面的 cross validation (cv=10) 去做 tuning 的。透過不斷的調整 n_estimators 以及 learning_rate 然後 tuning，來得到最好的 model。

   我的 public score 表現最好達到 0.88120，但其 private score 並不是很高，只有 0.87176。
   (n_estimators=355, learning_rate=0.165, random_state=112)

   另外一個 model 的 public score = 0.88108，private score = 0.87200。
   (我選擇這個當作 best model，n_estimators=356, learning_rate=0.165, random_state=112)

4. Refer to math problems

(1)

Likelihood Function = $\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}[log(p(x_n | c_k)) + log\pi_k]$

$\Rightarrow$ Maximize with constraint $\sum_{k=1}^{K} \pi_k = 1$ and Lagrangian multiplier.

$$\Rightarrow L(\pi, \lambda) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}[log(p(x_n | c_k)) + log\pi_k] + \lambda(\sum_{k=1}^{K} c_k - 1)$$

$$\frac{\partial L(\pi, \lambda)}{\partial \pi_k} = \frac{1}{\pi_k}\sum_{n=1}^{N} t_n k + \lambda = 0$$

$$\Rightarrow \pi_k = \frac{-N_k}{\lambda}$$

$$\frac{\partial L(\pi, \lambda)}{\partial \lambda} = \sum_{k=1}^{K} \pi_k - 1 = 0$$

$$\Rightarrow \sum_{k=1}^{K} \pi_k = 1$$

$$\Rightarrow \sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \frac{-N_k}{\lambda} = 1$$

$$\Rightarrow \lambda = -N$$

$$\Rightarrow \pi_k = \frac{N_k}{-(-N)} = \frac{N_k}{N}$$

(2)

$$\frac{\partial log(det(\Sigma))}{\partial \sigma_{ij}}$$

$$= \frac{1}{det(\Sigma)} \frac{\partial det(\Sigma)}{\partial \sigma_{ij}}$$

$$= \frac{1}{det(\Sigma)} (-1)^{i+j} M_{ij}$$

$$e_j \Sigma^{-1} e_i^T$$

$$= e_j \frac{\tilde{\Sigma}}{det(\Sigma)} e_i^T$$

$$= \frac{1}{det(\Sigma)} (-1)^{i+j} M_{ij}$$

( 本來是 $M_{ji}$，但是 $\tilde{\Sigma}$ 根據定義已經是 transpose 過的矩陣，故這邊 transpose 回來就是 $M_{ij}$ )

左式 = 右式，故得證。

(3)

**About** $\mu_k$：

$$\frac{\partial l(\mu, \Sigma | x^n)}{\partial \mu_k}$$

$$= \sum_{n=1}^{N} t_{nk} \Sigma^{-1} (\mu_k - x^n) = 0$$

$$\Rightarrow N_k \mu_k = \sum_{n=1}^{N} t_{nk} x^n \ (\Sigma^{-1} \ is \ positive \ infinite)$$

$$\Rightarrow \mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk}x^n$$

**About** $\Sigma$ :

$$l(\mu, \Sigma \mid x^n)$$

$$= C - \frac{N}{2}log\,|\,\Sigma\,| - \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}[t_{nk}(x^n - \mu_k)^T\Sigma^{-1}(x^n - \mu_k)]$$

$$= C + \frac{N}{2}log\,|\,\Sigma\,|^{-1} - \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}tr[t_{nk}(x^n - \mu_k)(x^n - \mu_k)^T\Sigma^{-1}]$$

$$\frac{\partial l(\mu, \Sigma \mid x^n)}{\partial \Sigma^{-1}}$$

$$= \frac{N}{2}\Sigma - \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n - \mu_k)(x_n - \mu_k)^T = 0,\ since\ \Sigma^T = \Sigma$$

$$\Rightarrow \Sigma = \frac{1}{N}\frac{N_k}{N_k}\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n - \mu_k)(x_n - \mu_k)^T$$

$$\Rightarrow \Sigma = \sum_{k=1}^{K}\frac{N_k}{N}\sum_{n=1}^{N}\frac{1}{N_k}t_{nk}(x_n - \mu_k)(x_n - \mu_k)^T$$

$$\Rightarrow \Sigma = \sum_{k=1}^{K}\frac{N_k}{N}s_k$$

$$where\ s_k = \frac{1}{N_k}\sum_{n=1}^{N}t_{nk}(x_n - \mu_k)(x_n - \mu_k)^T$$