學號：R07941023 系級： 光電碩二 姓名：呂彥穎

1. (0.5%) 請比較你實作的 generative model、logistic regression 的準確率，何者較佳?
Sol:
Kaggle (logistic regression):0.85429/0.84817
Kaggle (generative model):0.84324/0.84375
logistic regression 的效果比較好，其原因 generative model 是一個 data 彼此獨立的機率模型，在 data 少的時候具有一定程度的腦補想像力而準確度上升，但當 data 數量夠大的時候，logistic regression 減少 bias 帶來得誤差後，其準確度會大於 generative model。

2. (0.5%) 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
標準化的目的是讓多維的 feature 是個圓形，可以快一點到達 local minimum，如果時間是無限的，準確率不應該受到影響。
固定所有參數，只考慮 epochs:100 (注意:兩個都還沒收斂，僅只比較而已)
Original feature
Kaggle (logistic regression):0.23525/0.23719 (全都還是 1，收斂超慢)
Feature normalization
Kaggle (logistic regression):0.8371/0.8349

3. (1%) 請說明你實作的 best model，其訓練方式和準確率為何?
Sol:
Kaggle (best model):0.86953/0.85468
這裡使用的是 gradientboosting。
訓練步驟:
1.調個我破電腦不會跑太慢的 learning rate，這裡先設個 0.1
2.寫個迴圈去跑最低 loss 需要的”樹”數量，跑完後選擇 400 顆。
3.重複步驟 2，去寫迴圈單變數測試 depth/split/left/subsample
4.經過一番努力後，得到這些參數 depth = 8/split=50 /left=7/subsample=0.8
5.可靠資料說，這時候就可以降低 larning rate10 倍，然後讓樹增加 10 倍，結果壞掉了，所以回到上一步就直接跑分。

4. (3%) Refer to math problem

Sol:

(P1)

$$L(\pi,\lambda) = I(\theta) + \lambda\left(\sum_{k=1}^{k}\pi_k - 1\right), \quad I(\theta) = \sum_{n=1}^{k}\sum_{k=1}^{k} y_{n,k}\log(\ell(\theta))$$

$$= \sum_{n=1}^{k}\sum_{k=1}^{k} y_{n,k}\log \ell(\theta) + \lambda\left(\sum_{k=1}^{k}\pi_k - 1\right), \quad \ell(\theta) = \prod_{n=1}^{N}\prod_{k=1}^{k} \left(p(x_n|C_k)\pi_k\right)^{y_{n,k}}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{k} y_{n,k}\left[\log(P(x_n|C_k)) + \log\pi_k\right] + \lambda\left(\sum_{k=1}^{k}\pi_k - 1\right).$$

$$\frac{\partial L(\pi,\lambda)}{\partial \pi_k} = \frac{1}{\pi_k}\sum_{n=1}^{N} y_{n,k} + \lambda = 0, \quad \Rightarrow \pi_k = -\frac{1}{\lambda}\sum_{n=1}^{N} y_{n,k} = -\frac{N_k}{\lambda}$$

\#

(P2).

(13.44)

$$\frac{\partial\log(\det\Sigma)}{\partial\sigma_{ij}} = e_j\frac{1}{|\Sigma|}\frac{\partial}{\partial\sigma_{ij}}|\Sigma| e_i^T, \quad \overset{(13.46)}{\frac{\partial}{\partial\sigma_{ij}}|\Sigma| = \tilde{\Sigma}}$$

$$= e_j\frac{1}{|\Sigma|}\tilde{\Sigma} e_i^T, \quad \overset{(13.45)}{\Sigma^{-1} = \frac{1}{|\Sigma|}\tilde{\Sigma}}$$

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

\#

(P3) pf) $\mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk} x_n$

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

$\mu_k$

$$\frac{\partial P(x|C_k)}{\partial \mu_k} = 0 \quad , \quad P(x|C_k) = N(x|\mu_k, \Sigma)$$

$$\Rightarrow \frac{\partial N(x|\mu_k,\Sigma)}{\partial \mu_k} = 0 .$$

$$\Rightarrow \sum_{n=1}^{N_k} \Sigma^{-1} (\mu_k - x_n) = 0$$

$$\Rightarrow N_k \mu_k = \sum_{n=1}^{N} t_{nk} x_n \quad , \quad \mu_k = \frac{1}{N_k}\sum_{n=1}^{N} t_{nk} x_n \quad \#$$

$\Sigma$

$$\frac{\partial P(x|C_k)}{\partial \Sigma^{-1}} = 0 \quad , \quad P(x|C_k) = N(x|\mu_k, \Sigma)$$

$$\Rightarrow \frac{\partial N(x|\mu_k,\Sigma)}{\partial \Sigma^{-1}} = 0 \quad , \quad N(x|\mu_k \Sigma) = C - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N} (x_i - \mu_k)^T \Sigma^{-1} (x_i - \mu_k)$$

$$\overset{tr}{=} C + \frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{N} tr[(x_i - \mu_k)(x_i - \mu_k)^T \Sigma^{-1}]$$

$$\Rightarrow \frac{\partial}{\partial \Sigma^{-1}}\left( C + \frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{i=1}^{N} tr\left[(x_i - \mu_k)(x_i - \mu_k)^T \Sigma^{-1}\right]\right) = 0 \quad ; \quad \begin{cases} \frac{\partial}{\partial \Sigma^{-1}} \log|\Sigma^{-1}| = \Sigma^T \\ \frac{\partial}{\partial \Sigma^{-1}} tr(\Sigma^{-1} X) = X^T \end{cases}$$

$$\Rightarrow \frac{N}{2}\Sigma^T - \frac{1}{2}\sum_{i=1}^{N} (x_i - \mu_k)(x_i - \mu_k)^T = 0 \quad , \quad \Sigma = \Sigma^T$$

$$\Rightarrow \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N} (x_i - \mu_k)(x_i - \mu_k)^T = 0$$

$$\Rightarrow \Sigma = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu_k)(x_i - \mu_k)^T$$

加 手寫的 $S_k$ 是我的 $\Sigma$

$$\Sigma = \sum_{k=1}^{k} \frac{N_k}{N} \Sigma \quad \#$$