學號：R07922163 系級： 資工所二 姓名：徐浩翔

1. (0.5%) 請比較你實作的generative model、logistic regression 的準確率，何者較佳?

使用 logistic regression 做出來的效果比較好，下表為在 kaggle hw2 上的表現

	private	public
logistic regression	0.84989	0.85552
generative model	0.84387	0.84373

其中不管是公開或是隱藏的測試資料皆為

logistic regression比較好

而我上傳了一份全部預測零的csv檔到kaggle hw2

在public 及private 上分別得到

0.76474與0.76282的預測正確率

表示資料大部分都屬於第零類(收入不超過美金50K)

因此generative model會受到先驗機率影響，得出來的預測值自然會比較小，所以表現比logistic regression差

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

在以logistic regression當作模型的狀況下

沒有標準化的資料在經過exponential時會overflow

而在kaggle hw2上的預測正確率只有0.78599

只比預測全部都是零的高0.02左右

因為若沒有經過normalization有些feature的數值範圍是1000~2000有些是0~10

這樣在trainning時會給數值範圍比較小的有比較高的權重

讓這個feature dominate預測結果但不應該這樣子所以會

讓模型的準確率降很多

3. (1%) 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 sklearn ensemble 中的gradient boosting classifier

其中跑得最的好的一些參數為:

Gradient Boosting Classifier(n_estimators =1000 ,validation_fraction=0.01,n_iter_no_change=50)

將1%的資料做為validation set 防止model overfitting

超過50次迭代驗證集的loss沒下降就early stopping

以下是使用 gradient boosting classifier在kaggle hw2上的結果

private	publice
0.86500	0.87272
0.87274	0.87936
0.87274	0.88083
0.87274	0.88132(publice最高的正確率)
0.87286	0.87960
0.87372(比private第一高但未選擇)	0.87936

另外我也有試過做五次gradient boosting 然後將每次結果做voting 三個以上的model預測1就將值設為一否則為0

結果在public上會比單一model多預測錯一到兩筆

但在private上得到了leaderboard第四名的成績

其中public 更有第一高的正確率。

4. (3%) Refer to math problem

https://hackmd.io/0fDimqO7RaSCPpD_minSGQ?both

1.

$$\arg\max_{\pi_k} P(x_1|C_k)\,P(x_2|C_k)\cdots P(x_N|C_o)$$ WLOG $x_1 \cdots x_{N_k} \in C_k$

C_o 為 other class

$$\arg\max_{\pi_k} \prod_{i=1}^{N_k} \frac{P(x_i, C_k)}{P(C_k)} \prod_{j=N_k+1}^{N} \frac{P(x_j, C_o)}{P(C_o)}$$

$$\Rightarrow \arg\max_{\pi_k} \left(\frac{\frac{1}{N_k}}{\pi_k}\right)^{N_k} \left(\frac{\frac{1}{N-N_k}}{1-\pi_k}\right)^{N-N_k}$$

取 ln

$$\Rightarrow \arg\max_{\pi_k} N_k \ln \frac{1}{N_k \pi_k} + (N-N_k) \ln \frac{1}{(N-N_k)(1-\pi_k)}$$

$$\Rightarrow \arg\max_{\pi_k} -N_k \ln N_k \pi_k - (N-N_k) \ln (N-N_k)(1-\pi_k)$$

$$\frac{\partial}{\partial \pi_k} \Rightarrow -N_k \frac{1}{\pi_k} - (N-N_k) \frac{1}{1-\pi_k}(-1) = 0$$

$$\Rightarrow -\frac{N_k}{\pi_k} + \frac{N-N_k}{1-\pi_k} = 0$$

$$\Rightarrow \frac{N_k}{\pi_k} = \frac{N-N_k}{1-\pi_k} \Rightarrow N_k - N_k\pi_k = N\pi_k - N_k\pi_k$$

$$\Rightarrow N_k = N\pi_k$$

$$\pi_k = \frac{N_k}{N}$$ ※

2.

$$\det\Sigma = \det\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & & & \\ \sigma_{31} & & \ddots & \\ \vdots & & & \\ \sigma_{m1} & \cdots & & \sigma_{mm} \end{bmatrix}$$

$= \sigma_{11}\det C_{11} + \sigma_{12}C_{12} \cdots \sigma_{1m}C_{1m}$ 其中 C_{1m} 為 cofactor matrix

(C_{ij} 為 Σ 扣掉 i 列 j 行所得的 $m-1 \times m-1$ 矩陣)

$= \sigma_{i1}\det C_{i1} + \sigma_{i2}C_{i2} \cdots + \sigma_{ij}C_{ij} + \cdots \sigma_{im}C_{im}$

$$\frac{\partial \log(\det\Sigma)}{\partial_{ij}} = \frac{1}{\det\Sigma}C_{ij}$$

故題目左式為 $\frac{1}{\det\Sigma}C_{ij}$

classical adjoint matrix $\mathrm{adj}(A) = C^T$ $C = \begin{bmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mm} \end{bmatrix}$

$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A)$ 而 $e_j\Sigma^{-1}e_i^T = \Sigma^{-1}_{ji}$ (題目右式)

$$\downarrow$$

$$\Sigma^{-1}_{ji} = \frac{1}{\det(\Sigma)}\underbrace{C_{ij}}_{(\because \mathrm{adj}\,A = C^T)}$$

3-(a)

$L(\mu_k,\Sigma|x) = P(x_1|C_k)P(x_2|C_k)\cdots P(x_{n_k}|C_k)$ 其中 WLOG $t_{1k}\, t_{2k} \cdots t_{n_k k} = 1$

likelihood

$$= N(x_1|\mu_k,\Sigma)N(x_2|\mu_k,\Sigma)\cdots N(x_{n_k}|\mu_k,\Sigma)$$

$$= \left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}e^{-\frac{(x_1-\mu_k)^T\Sigma^{-1}(x_1-\mu_k)}{2}}\right)\cdots\cdots\left(\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\Sigma|^{\frac{1}{2}}}e^{-\frac{(x_{n_k}-\mu_k)^T\Sigma^{-1}(x_{n_k}-\mu_k)}{2}}\right)$$

取 ln

$$\ln(L(\mu_k,\Sigma|x)) = \left(-\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_1-\mu_k)^T\Sigma^{-1}(x_1-\mu_k)\right)+\cdots+\left(-\frac{D}{2}\ln 2\pi + \frac{1}{2}(x_{n_k}-\mu_k)^T\Sigma^{-1}(x_{n_k}-\mu_k)\right)$$

$$\frac{\partial \ln(L(\mu_k,\Sigma|x))}{\partial \mu_k} = \frac{1}{2}\left(-\Sigma^{-1}(x_1-\mu_k)-\left((x_1-\mu_k)^T\Sigma^{-1}\right)^T\right)+\cdots+\frac{1}{2}\left(-\Sigma^{-1}(x_{n_k}-\mu_k)-\left((x_{n_k}-\mu_k)^T\Sigma^{-1}\right)^T\right)$$

$$= -\Sigma^{-1}(x_1-\mu_k)-\Sigma^{-1}(x_2-\mu_k)\cdots-\Sigma^{-1}(x_n-\mu_k) = \vec{0}$$

$\because \Sigma^{-1}$ is covariance matrix is symmetric and nonsingular

$\Rightarrow \quad \Sigma^{-1}x = \vec{0} \quad \Leftrightarrow \quad \vec{x} = \vec{0}$

故 $\sum_{i=1}^{n_k}(x_i-\mu_k) = 0$

$$\mu_k = \frac{\sum_{i=1}^{n_k}x_i}{n_k} = \frac{1}{n_k}\sum_{i=1}^{N}t_{ik}x_i \quad (\because t_{ik}x_i = 0 \text{ if } i > n_k)$$

(n_k為題目N_k符號來不及改)

3-(b)

similarly to (a)

$$\ln(L(\mu_k,\Sigma|X^k)) = \overbrace{\left(-\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_1-\mu_k)^T\Sigma^{-1}(x_1-\mu_k)\right)}^{x_1 \cdots x_{n_k} \in C_k} + \cdots\cdots$$

$$+\left(-\frac{D}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x_{n_k}-\mu_k)^T\Sigma^{-1}(x_{n_k}-\mu_k)\right)$$

$$\boxed{tr([\,][\,]) = tr([\,][\,])}$$

$$\Rightarrow \ln(L(\mu_k,\Sigma|X)) = \left(-\frac{D}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\left(tr\left((x_1-\mu_k)(x_1-\mu_k)^T\Sigma^{-1}\right)\right)\right) \cdots\cdots$$

$$+\left(-\frac{D}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\left(tr\left((x_{n_k}-\mu_k)(x_{n_k}-\mu_k)^T\Sigma^{-1}\right)\right)\right)$$

$$\frac{\partial \ln(L(\mu_k,\Sigma|X))}{\partial \Sigma^{-1}} = \frac{n_k}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n_k}(x_i-\mu_k)(x_i-\mu_k)^T$$

$\left(\because \frac{\partial \ln|\Sigma|}{\partial \Sigma} = (\Sigma^{-1})^T\right.$ by problem and Σ is covariance matrix is symmetric$\left.\right)$

$$\Rightarrow \Sigma = S_k$$

share Σ so need to find

$$L(\mu_1,\Sigma|X^1)L(\mu_2,\Sigma|X^2)\cdots L(\mu_k,\Sigma|X^k)$$ and $X^1, X^2 \cdots X^k$

$\forall x_i \in X^1 \quad x_i \in C_k$

$\Rightarrow$ 取 $\ln$ 再 $\partial\Sigma^{-1}$

WLOG X^1為$(x_1 \cdots x_{n_k})$

$$\Rightarrow \left(\frac{n_1}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n_1}(x_i-\mu_1)(x_i-\mu_1)^T\right) + \cdots - \left(\frac{n_k}{2}\Sigma - \frac{1}{2}\sum_{i=N-n_k}^{N}(x_i-\mu_k)(x_i-\mu_k)^T\right) = 0$$

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