學號：r08942085 系級： 電機四 姓名：陳芃彣

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

	private	public
generative_v1.csv 3 days ago by r08942085 add submission details	0.84363	0.84410
logistic_v1.csv 3 days ago by r08942085 add submission details	0.84915	0.85429

logistic regression在public與private的成績都比generative model佳，雖然分數很接近但logistic regression的準確率高一些。

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

Submission and Description	Private Score	Public Score
best_test_v1.csv a few seconds ago by r08942085 add submission details	0.86574	0.87714
best_random_forest_v8.csv 4 days ago by r08942085 add submission details	0.85763	0.87051

上圖的best_test_v2.csv以及best_rangdom_forest_v8.csv皆是使用gradientboostingclassifier，上者為資料未正規化的，下者為有資料更規劃的，然而出來的結果有點不符合我的預期，本來預期正規劃後的成績應該會比較高，因為正歸化後可以減少資料的差異性，但卻有不一樣的結果。

Submission and Description	Private Score	Public Score
logistic_test_v1.csv just now by r08942085 add submission details	0.30733	0.30307
logistic_v1.csv 3 days ago by r08942085 add submission details	0.84915	0.85429

但使用logistic regression就有明顯的差異，為正規化的資料使用logistic regression出來的成績只有0.3。

Submission and Description	Private Score	Public Score
generative_test_v1.csv just now by r08942085 add submission details	0.84412	0.84385
generative_v1.csv 3 days ago by r08942085 add submission details	0.84363	0.84410

generative model正規劃前與後的差異不大，未正規劃的public score低於有正規化的一些，然private score卻高了一些，不過兩者分數是很接近的。

3. (1%) 請說明你實作的**best model**，其訓練方式和準確率為何?

我的best model使用的是scikit-learn的GradientBoostingClassifier(n_estimator=75、max_depth=9、randomstate=0)，準確率為private: 0.85763，public: 0.87051。

（圖中的輸出檔案名字取錯了，model是試使用gradientboosting）

best_random_forest_v8.csv 4 days ago by r08942085 add submission details	0.85763	0.87051

4. (3%) Refer to math problem

陳芃彣 R08942085

1. $p(C_k) = \pi_k \quad k=1,2\cdots K$

$p(x|C_k) \quad k=1,2\ldots K$

$t = [t_1, t_2 \cdots t_K]$

$= [0,1,0\cdots 0]$

The probability of one data point:

$$P(x,t) = P(x|t)P(t) = (P(x|C_1)\pi_1)^{t_1}(P(x|C_2)\pi_2)^{t_2}\cdots(P(x|C_K)\pi_K)^{t_K}$$

$$= \prod_{k=1}^{K}(P(x|C_k)\pi_k)^{t_k}$$

令 parameter of model $= \theta$

$$L(\theta) = \prod_{n=1}^{N}\prod_{k=1}^{K}(P(x_n|C_k)\pi_k)^{t_{n,k}}$$

n筆資料

取log (相乘變相加)

$$l(\theta) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}[\log P(x_n|C_k) + \log \pi_k]$$

log likelihood s.t. $\sum_{k=1}^{K}\pi_k = 1$

$g(\pi,\lambda)=0$

$\Rightarrow$ lagrange multiplier:

$$\mathcal{L}(\pi,\lambda) = \sum_{n=1}^{N}\sum_{k=1}^{K} t_{n,k}[\log P(x_n|C_k) + \log\pi_k] + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)$$

$$\begin{cases} \dfrac{\partial}{\partial \pi_k}\mathcal{L} = \dfrac{1}{\pi_k}\displaystyle\sum_{n=1}^{N} t_{n,k} + \lambda = 0 \Rightarrow \pi_k = -\dfrac{1}{\lambda}\displaystyle\sum_{n=1}^{N} t_{n,k} = -\dfrac{N_k}{\lambda} \quad \text{①} \\ \dfrac{\partial}{\partial \lambda}\mathcal{L} = \displaystyle\sum_{k=1}^{K}\pi_k - 1 \Rightarrow \sum_{k=1}^{K}\pi_k = 1 \end{cases}$$

$$\sum_{k=1}^{K}\pi_k = \sum_{k=1}^{K}\frac{-N_k}{\lambda} = \frac{-N}{\lambda} = 1 \Rightarrow \lambda = -N \text{ 代入① } \rightarrow \pi_k = \frac{N_k}{N} \#$$

①

2. $$\frac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = e_j \Sigma^{-1} e_i^T$$

where $\Sigma \in \mathbb{R}^{m\times m}$ is a covariance matrix and e_j is a row vector (ex: $e_3 = [0,0,1,\cdots,0]$)

令 $X^{-1} = \Sigma$

$$\log \det X^{-1} = \log(\det X)^{-1} = -\log \det X$$

$$\frac{\partial}{\partial X_{ij}} \log\det X^{-1} = -\frac{\partial}{\partial X_{ij}} \log \det X = -\frac{1}{\det X}\frac{\partial \det X}{\partial X_{ij}} = -\frac{1}{\det X} \operatorname{adj}(X_{ji})$$

$$= -(X^{-1})_{ji} \quad \#$$

3. problem 1 $\rightarrow p(C_k) = \pi_k \ , \ k = 1, 2 \cdots K$

problem 2 $\rightarrow \dfrac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = e_j \Sigma^{-1} e_i^T$

$$p(x \mid C_k) = \mathcal{N}(x \mid \mu_k, \Sigma)$$

$\rightarrow \mu_k = \dfrac{1}{N_k} \sum_{n=1}^{N} t_{nk} x_n$ for C_k

$\rightarrow \Sigma = \sum_{k=1}^{K} \dfrac{N_k}{N} S_k$ where $S_k = \dfrac{1}{N_k} \sum_{n=1}^{N} t_{nk} (x_n - \mu_k)(x_n - \mu_k)^T$

$$L(\mu, \Sigma) = f_{\mu,\Sigma}(x^1) f_{\mu,\Sigma}(x^2) \cdots f_{\mu,\Sigma}(x^N)$$

$$= \prod_{n=1}^{N} f_{\mu,\Sigma}(x^n)$$

取log

$$l(\mu, \Sigma \mid x^{(n)}) = \log \prod_{n=1}^{N} f(x^{(n)} \mid \mu, \Sigma)$$

$$= \log \prod_{n=1}^{N} \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu)\right)$$

$$= \sum_{n=1}^{N} \left(\frac{-p}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu) \right)$$

$$= \frac{-Np}{2} \log(2\pi) - \frac{N}{2} \log|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} \left[(x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu) \right]$$

$\downarrow$

要找maximum $\rightarrow$ 微分求極值

$$\frac{\partial l}{\partial \mu_k} = \sum_{n=1}^{N} \Sigma^{-1} (\mu_k - x^{(n)}) = 0$$

$\dfrac{\partial w^T A w}{\partial w} = 2Aw$

Assume $\underbrace{(x^{(n)} - \mu)}_{w}$ does not depend on $\underbrace{\Sigma^{-1}}_{A}$ & Σ^{-1} is symmetric

Since Σ is positive definite $\rightarrow$ 每一element皆為正

$$0 = N_k \mu_k - \sum_{n=1}^{N} t_{nk} x_n \longleftarrow \therefore \sum_{n=1}^{N} (\mu_k - x^{(n)}) = 0$$

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

$$l(\mu, \Sigma \mid x^{(n)}) = C - \frac{N}{2} \log|\Sigma| - \frac{1}{2} \sum_{n=1}^{N} \left[(x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu) \right]$$

$\because x^t A x = tr(x^T A x) = tr(x^t x A)$

$tr(ABC) = tr(CAB) = tr(BCA)$

$$= C + \frac{N}{2} \log|\Sigma^{-1}| - \frac{1}{2} \sum_{n=1}^{N} tr\left[(x^{(n)} - \mu)(x^{(n)} - \mu)^T \Sigma^{-1} \right]$$

$$\frac{\partial}{\partial \Sigma^{-1}} l(\mu, \Sigma \mid x^{(n)}) = \frac{N}{2} \Sigma - \frac{1}{2} \sum_{n=1}^{N} (x^{(n)} - \mu)(x^{(n)} - \mu)^T = 0 \quad \text{Since } \Sigma^T = \Sigma$$

$$0 = N\Sigma - \sum_{n=1}^{N} (x^{(n)} - \mu)(x^{(n)} - \mu)^T$$

$$\Sigma = \frac{1}{N} \sum_{n=1}^{N} (x^{(n)} - \mu)(x^{(n)} - \mu)^T$$

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

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