學號：B07901169　　系級：　電機三　姓名：楊宗桓

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

	train	validation	public	private
generative model	0.83532	0.83355	0.84508	0.83269
logistic regression	0.85006	0.84154	0.84791	0.84289

兩種 model 都是經過 normalized ，validation 跟 training set 為將 training data 切一半得來，接著去除'fnlwgt'，因為我上網查發現 fnlwgt 是代表資料中某個州的權重人數，對收入判斷沒有幫助，所以刪除。

如上圖: logistic regression 不論在 private 或 public set 都有較高的準確率，這是因為 generative model 在一開始就假設資料有某種特定的分布，但實際並非如此。

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

		training	validation	public	private
unnormalized	generative model	0.83581	0.83379	0.84484	0.83233
	logistic regression	0.83031	0.83259	0.82877	0.82387
normalized	generative model	0.83532	0.83355	0.84508	0.83269
	logistic regression	0.85006	0.84154	0.84791	0.84289

Normalized 的部分只針對 X_train 中 age,capital_gain,capital_loss,hours_per_week 因為我發現只有這幾項是沒有經過 one-hot encodeing 且其值為連續值。

可以發現 normalization 對 generative model 影響不大，但對 logistic regression 影響較大，但不論是哪種模型，做了 normalization 都可使模型的準確度增加。

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 是使用 sk-learn 的 GradientBoostingClassifier，其中我發現將 learning_rate 調成 0.05，n_estimators = 800 可以得到較高的準確度，最後的準確率為:

train	validation	public	private
0.88401	0.872	0.87506	0.86991

我在 kaggle 上選的 best_model(lr=0.05，n_estimators = 500 準確率為:

train	validation	public	private
0.8769	0.87009	0.87432	0.8677

可以發現一件有趣的事，用這個套件做出來的模型，只要把訓練次數一直提高準確率一定會跟著提升(除非達到最小值)而不會 overfit。

4. (3%) Refer to math problem

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

ML HW #2.

1.

Likelihood function:

$$P(x_1, x_2, \dots x_N) = P(x_1) \times P(x_2) \times P(x_3) \dots P(x_N)$$

let C_{x_i} denote the class x_i belongs to.

$$\rightarrow \prod_{i=1}^{N} P(C_{x_i}) P(x_i \mid C_{x_i})$$

同取 log

$$\log P = \sum_{i=1}^{N} \log P(C_{x_i}) + \sum_{i=1}^{N} \log P(x_i \mid C_{x_i})$$

$\because$ N_k denotes the number of datas that belongs to C_k

$$\sum_{i=1}^{N} \log P(C_{x_i}) = \sum_{i=1}^{K} N_i \log P(C_i) = \sum_{i=1}^{K} N_i \log \pi_i$$

To find π_i that Maximize $P(x_1, x_2, x_3 \dots x_N)$

$$\frac{\partial \log P}{\partial \pi_i} = \frac{\partial}{\partial \pi_i}\left(\sum_{i=1}^{K} N_i \log \pi_i + \sum_{i=1}^{N} \log P(x_n \mid C_{x_n}) \right)$$

$\because$ datas are drawn independently from the model $\Rightarrow$ independent of π_i

$$= \frac{\partial}{\partial \pi_i} \sum_{i=1}^{K} N_i \log \pi_i$$

$$= \frac{\partial}{\partial \pi_i} N_i \log \pi_i = \frac{N_i}{\pi_i}$$

Because $\sum \pi_i = 1$, from Lagrange Multiplier we know the maximum of $\log P$ (ie P) happens when $\nabla f(x,y) = \lambda \nabla (g(x,y) - c)$

where $f = \log P$; $g = \sum \pi_i$, $C = 1$, $\frac{\partial g}{\partial \pi_i} = 1$

$$\Rightarrow \frac{N_i}{\pi_i} = \lambda \qquad \sum_{i=1}^{N} \pi_i = 1 \Rightarrow \frac{N}{\lambda} = 1, \ \lambda = N$$

$$\Rightarrow \pi_i = \frac{N_i}{\lambda} = \frac{N_i}{N} \quad \#$$

2. $\boxed{\text{Ref: wikipedia: 餘因子矩陣}}$

To compute det of $n \times n$ Matrix using Cofactor Expansion:

$$\det(A) = (-1)^{i+1} A_{i,1} \det(A(i|1)) + (-1)^{i+2} A_{i,2} \det(A(i|2)) + \dots + (-1)^{i+n} A_{i,n} \det(A(i|n))$$

where A: $n \times n$ Matrix, let C_{ij} denote $(-1)^{i+j} \det(A(i|j))$

(1)

$$\Rightarrow \det(A) = \sum_{j=1}^{n} A_{i,j} C_{i,j} = \sum_{i=1}^{n} A_{i,j} C_{i,j} \quad (2)$$

Back to the problem:

By (1)

$$\det \Sigma = \sum_{j=1}^{n} \sigma_{i,j} C_{ij}$$

$$\frac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = \frac{1}{\det \Sigma} \frac{\partial}{\partial \sigma_{ij}} \left(\sum_{j=1}^{n} \sigma_{ij} C_{ij} \right)$$

by (2).

$$= \frac{C_{ij}}{\det \Sigma} \quad \text{where } C_{ij} = (\operatorname{adj} \Sigma)_{j,i}$$

$$= \left(\frac{\operatorname{adj} \Sigma}{\det \Sigma} \right)_{ji}$$

by (3).

$$= (\Sigma^{-1})_{ji} = e_j \Sigma^{-1} e_i^T$$

#

$$\operatorname{Adj}(A) = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ \vdots & & & \\ C_{1n} & & \cdots & C_{nn} \end{bmatrix}$$

: transpose of cofactor Matrix.

(3)

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

3.

Now, we need to find the mean, covariance Matrix to maximize $P(x_1, x_2 \dots x_n)$ from problem 1, this equals to find maximum of $\log P$.

$$\log P = \sum_{i=1}^{K} N_i \log \pi_i + \sum_{i=1}^{N} \log P(x_i | C_{x_i})$$

$$\sum_{i=1}^{K} \log P(x_i | C_{x_i}) = \sum_{i=1}^{K} \sum_{x \in C_i} \log P(x | C_i)$$

$\because t_{nk} = 1$ when $x \in C_k$, otherwise $= 0$

$$= \sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \log P(x_n | C_i)$$

從題意 $p(x | C_k) = N(x | u_k, \Sigma)$

$$= \sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \log N(x_n | u_i, \Sigma) = f(u, \Sigma)$$

$\Rightarrow$ a variable of u, Σ, find u, Σ to maximize f equals to finding maximum of P.

$$\frac{\partial f}{\partial u_i} = \frac{\partial}{\partial u_i} \sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \log N(x_n | u_i, \Sigma)$$

$$N(x_n | u_i, \Sigma) = \frac{1}{\left((2\pi)^{m} \det(\Sigma)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n - u_i)^T \Sigma^{-1} (x_n - u_i)}$$

$$\log N = -\frac{1}{2}(x_n - u_i)^T \Sigma^{-1} (x_n - u_i) - \frac{m}{2}\log 2\pi - \frac{1}{2}\log \det \Sigma$$

function of Σ, indep. of u.

$$\frac{\partial \log N}{\partial u_i} = -\frac{1}{2}\frac{\partial}{\partial u_i}\left((x_n - u_i)^T \Sigma^{-1} (x_n - u_i)\right)$$

$$= -\frac{1}{2}\left[(-1)\Sigma^{-1}(x_n - u_i) + (x_n - u_i)^T \Sigma^{-1}(-1)\right)$$

$$= \Sigma^{-1}(x_n - u_i)$$

$$\frac{\partial f}{\partial u_i} = \sum_{n=1}^{N} t_{ni} \frac{\partial}{\partial u_i} \log N = \sum_{n=1}^{N} t_{ni} \Sigma^{-1}(x_n - u_i)$$

3. 續

$$\sum_{n=1}^{N} t_{ni} \Sigma^{-1} (x_n - u_i)$$

$$= \sum_{n=1}^{N} \Sigma^{-1} t_{ni} x_n - \sum_{n=1}^{N} \Sigma^{-1} t_{ni} u_i = \Sigma^{-1} \left(\sum_{n=1}^{N} (t_{ni} x_n) - N_i u_i \right)$$

to let this $= 0$

$$\sum t_{ni} x_n - N_i u_i = 0$$

$$u_i = \frac{1}{N_i} \sum_{n=1}^{N} t_{ni} x_n$$

#

$$\sum_{n=1}^{N} t_{ni} = N_i$$

if $x_n \in C_i \Rightarrow t_{ni} = 1$

$x_n \notin C_i \Rightarrow t_{ni} = 0$

$\therefore \sum_{n=1}^{N} t_{ni}$ = number of datas belonging to class i

Now, we find Σ that maximize f:

$$\frac{\partial f}{\partial \Sigma} = \frac{\partial}{\partial \Sigma} \sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \log N(x_n | u_i, \Sigma)$$

$$= \sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \frac{\partial}{\partial \Sigma} \log N(x_n | u_i, \Sigma)$$

$$\log N = -\frac{1}{2} (x_n - u_i)^T \Sigma^{-1} (x_n - u_i) - \frac{1}{2} \log \det \Sigma - \frac{m}{2} \log 2\pi$$

from Problem 2, $\frac{\partial \log \det \Sigma^{-1}}{\partial \sigma_{ij}} = \Sigma^{-1}_{ji}$

$$\Rightarrow \frac{\partial \log \det \Sigma}{\partial \Sigma} = (\Sigma^{-1})^T$$

$$\frac{\partial \log N}{\partial \Sigma} = -\frac{1}{2} (x_n - u_i)^T \frac{\partial \Sigma^{-1}}{\partial \Sigma} (x_n - u_i) - \frac{1}{2} (\Sigma^{-1})^T$$

$$= -\frac{1}{2} (x_n - u_i)^T (x_n - u_i) \cdot \Sigma^{-2}$$

$$\sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \left(\frac{1}{2} (x_n - u_i)^T (x_n - u_i) \Sigma^{-2} - \frac{1}{2} (\Sigma^{-1})^T \right)$$

① $\sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} (\Sigma^{-1})^T = N (\Sigma^{-1})^T$

② $\sum_{i=1}^{K} \sum_{n=1}^{N} t_{ni} \frac{1}{2} (x_n - u_i)^T (x_n - u_i) \Sigma^{-2}$

$$= \sum_{i=1}^{K} N_i S_i \Sigma^{-2}$$

To let $\frac{\partial \log N}{\partial \Sigma} = 0$

$$\frac{1}{2}\left(\sum_{k=1}^{K} N_k S_k \Sigma^{-2} - N\left(\Sigma^{-1}\right)^T\right) = 0$$

$$\sum_{k=1}^{K} N_k S_k \Sigma^{-2} = N\left(\Sigma^{-1}\right)^T$$

同乘 Σ^2

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

#