<u>ML Homework#2</u> 學號:B0902120 系級: 資工四 姓名:曾鈺婷

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

經過比較之後，我發現 logistic regression 的效果比較好。可能是因爲 sigmoid function 的使用，讓整體資料相對平滑化，比較沒有極值的出現；除此之外，總共116維的情況（加入特定 column 的二次及三次）之下，會讓 covarience matrix 的 varience 變大，這也可能是導致表現變差的原因之一。

	Public Score	Private
Logistic Regression	0.85773	0.85444
Probabilistic Generative	0.84434	0.84399

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。

這次我採取了兩種方式，一種是 rescaling（min 與 max）另一種則是 standardization，這兩種方式的效果其實沒有差很多，但相較於什麼都不做，他的效果還是明顯好很多，因此我們可以得知 normalization 是必要的，並且可以擇一就好。

	Public Score	Private
None	0.23525	0.23719
Rescaling	0.84275	0.83970
Standardization	0.85515	0.85296

3. 請說明你實作的best model，其訓練方式和準確率為何？

在這次的作業當中，我額外加上了一些比較重要 column 的二次及三次方，除此之外，根據 greedy algorithm 我採取的是 logistic regression model 並使用 standardization 去做 feature normalization。

4. 手寫題

1.

$$P(x_1, x_2, \ldots, x_N) = \prod_{n=1}^{N} P(x_n) = \prod_{n=1}^{N} P(C_{x_n}) P(x_n \mid C_{x_n})$$

$$\log P(x_1, x_2, \ldots, x_N) = \sum_{n=1}^{N} \log[P(C_{x_n}) P(x_n \mid C_{x_n})]$$

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

$$= \sum_{k=1}^{K} N_k \log P(C_k) + \sum_{n=1}^{N} \log P(x_n \mid C_{x_n})$$
$$= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{n=1}^{N} \log P(x_n \mid C_{x_n})$$

透過 Lagrange Multiplier 可以算出

$$\frac{\partial}{\partial \pi_i}[\log P(x_1, x_2, \ldots, x_N)] = \frac{\partial}{\partial \pi_i}[\sum_{k=1}^{K} N_k \log \pi_k + \sum_{n=1}^{N} \log P(x_n \mid C_{x_n})]$$
$$= \frac{\partial}{\partial \pi_i}[N_i \log \pi_i + 0] = \frac{N_i}{\pi_i}$$

$$\frac{\partial}{\partial \pi_i}[\sum_{k=1}^{K} \pi_k] = \frac{\partial}{\partial \pi_i}[p_i] = 1$$

夠過 $\frac{N_i}{\pi_i} = \lambda * 1$ 且 $\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \frac{N_k}{\lambda} = \frac{N}{\lambda} = 1$ 可得到 $\lambda = N$，帶入得到$\pi_i = \frac{N_i}{N}$

2.

$$\frac{\partial}{\partial \sigma_{ij}}[\log(det\Sigma)] = \frac{1}{det\Sigma}\frac{\partial}{\partial \sigma_{ij}}[det\Sigma]$$
$$= \frac{1}{det\Sigma}\frac{\partial}{\partial \sigma_{ij}}[\sigma_{i1}C_{i1} + \sigma_{i2}C_{i2} + \ldots + \sigma_{im}C_{im}] \ \text{(cofactor elements)}$$
$$= \frac{1}{det\Sigma}C_{ij}$$
$$= \frac{1}{det\Sigma}(adj\Sigma)_{ji}$$ (reference: https://zh.wikipedia.org/wiki/伴隨矩陣)
$$= (\Sigma^{-1})_{ji} = e_j \Sigma^{-1} e_i^T$$

3.

$$\log P(x_1, x_2, \ldots, x_N) = \sum_{k=1}^{K} N_k \log \pi_k + \sum_{n=1}^{N} \log P(x_n \mid C_{x_n})$$
$$= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{k=1}^{K} \sum_{x \in C_k} \log P(x \mid C_k)$$
$$= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} t_{nk} \log P(x_n \mid C_k) \ (t_{nk} = 1 \text{ only if } n \text{ is from } k)$$
$$= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} t_{nk} \log N(x_n \mid \mu_k, \Sigma)$$

$$
\begin{aligned}
&= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk} \log[\frac{1}{\sqrt{(2\pi)^m\ det\Sigma}} e^{-\frac{1}{2}(\mu_k - x_n)^T\Sigma^{-1}(\mu_k - x_n)}] \\
&= \sum_{k=1}^{K} N_k \log \pi_k + \sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}[-\frac{1}{2}(\mu_k - x_n)^T\Sigma^{-1}(\mu_k - x_n) - \frac{1}{2}\log det\Sigma - \frac{m}{2}\log 2\pi]
\end{aligned}
$$

只有後項與題目要求有關，因此我們也只需考慮後項的最大值。

$$
\begin{aligned}
&\frac{\partial}{\partial \mu_i}\sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}[-\frac{1}{2}(\mu_k - x_n)^T\Sigma^{-1}(\mu_k - x_n) - \frac{1}{2}\log det\Sigma - \frac{m}{2}\log 2\pi] \\
&= \sum_{n=1}^{N} t_{ni}[-\frac{1}{2} * 2 * \Sigma^{-1}(\mu_i - x_n)] \\
&= \Sigma^{-1}\sum_{n=1}^{N} t_{ni}[x_n - \mu_i] \\
&= \Sigma^{-1}(\sum_{n=1}^{N} t_{ni}x_n - N_i\mu_i)
\end{aligned}
$$

極值發生在微分為 0 的時候，因此我們可以得知：

$$
\begin{aligned}
&\sum_{n=1}^{N} t_{ni}x_n - N_i\mu_i = 0 \\
&\mu_i = \frac{1}{N_i}[\sum_{n=1}^{N} t_{ni}x_n]
\end{aligned}
$$

第二部分的證明，也是採取相似的方法。

$$
\begin{aligned}
&\frac{\partial}{\partial \Sigma^{-1}}\sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}[-\frac{1}{2}(\mu_k - x_n)^T\Sigma^{-1}(\mu_k - x_n) - \frac{1}{2}\log det\Sigma - \frac{m}{2}\log 2\pi] \\
&= \frac{\partial}{\partial \Sigma^{-1}}\sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}[-\frac{1}{2}(\mu_k - x_n)^T\Sigma^{-1}(\mu_k - x_n) - \frac{1}{2}\log \frac{1}{det\Sigma^{-1}} - \frac{m}{2}\log 2\pi] \\
&= \sum_{k=1}^{K}\sum_{n=1}^{N} t_{nk}[-\frac{1}{2}(\mu_k - x_n)^T(\mu_k - x_n) - \frac{1}{2}(-\Sigma)] \\
&= \frac{1}{2}\sum_{k=1}^{K}[\sum_{n=1}^{N} t_{nk}\Sigma - \sum_{n=1}^{N}(\mu_k - x_n)^T(\mu_k - x_n)] \\
&= \frac{1}{2}\sum_{k=1}^{K}[N_k\Sigma - N_kS_k] \\
&= \frac{1}{2}[N\Sigma - \sum_{k=1}^{K} N_kS_k] \\
&\Sigma = \frac{1}{N}[\sum_{k=1}^{K} N_kS_k]
\end{aligned}
$$