學號：R08725021 系級： 資管碩一 姓名：王鼎元

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

下表為兩個模型在 Kaggle 上的準確率：

	Public scores	Private scores
Generative model	0.81695	0.81390
Logistic regression	0.73461	0.72352

由上表可知，在沒有特徵標準化的情形下 Generative model 表好較好。

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

下表為在 epoch = 1000，batch size = 64，optimizer = Adam 的模型下 ，Logistic regression 在資料處理不同的情形下，在 Kaggle 上的準確率：

	Public scores	Private scores
Original data	0.76474	0.76280
Normalized data	0.85515	0.85026

由上表可知，特徵標準化後會讓模型更容易收斂，並且不會出現 model 運算時 overflow 的風險。

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 scikit-learn 的 SVC 模組來實作 SVM，kernel = RBF，利用 cross validation 所得到的 validation error 來 tune 參數 γ ，下表為在 Kaggle 上的準確率：

	Public scores	Private scores
SVM(gamma=0.2, C=1)	0.85909	0.85542

4. (3%) Refer to math problem

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

Q1.

1. Likelihood: $L(\theta) = \prod_{n=1}^{N} \prod_{k=1}^{K} [P(X_n|C_k) \cdot \pi_k]^{t_{nk}}$

→ log-likelihood: $\ell(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{nk} \cdot [\log P(X_n|C_k) + \log \pi_k]$

→ ∵ $\ell(\theta)$ subject to $\sum_{k=1}^{K} \pi_k = 1$

→ Lagrange multiplier:

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

$$\rightarrow \frac{\partial}{\partial \pi_k} \mathcal{L}(\pi, \lambda) = \frac{1}{\pi_k} \sum_{n=1}^{N} t_{nk} + \lambda = 0$$

$$\Rightarrow \because \sum_{n=1}^{N} t_{nk} = N_k \longrightarrow \pi_k = -\frac{N_k}{\lambda}$$

$$\rightarrow \frac{\partial}{\partial \lambda} \mathcal{L}(\pi, \lambda) = \sum_{k=1}^{K} \pi_k - 1$$

$$= \sum_{k=1}^{K} -\frac{N_k}{\lambda} - 1 = -\frac{N}{\lambda} - 1 = 0$$

$$\Rightarrow \lambda = -N$$

$$\Rightarrow \pi_k = \frac{N_k}{N}$$ #

Q2.

2. $$\frac{\partial}{\partial \sigma_{ij}} \log(\det \Sigma) = \frac{1}{\det \Sigma} \frac{\partial}{\partial \sigma_{ij}} (\det \Sigma)$$

$\longrightarrow$ by Jacobi's formula

$$\Rightarrow \frac{\partial}{\partial \sigma_{ij}} (\det \Sigma) = \tilde{\Sigma}_{ij}$$, $\tilde{\Sigma}$ is matrix of cofactor

$$\Rightarrow \frac{\partial}{\partial \sigma_{ij}} \log(\det \Sigma) = \frac{1}{\det \Sigma} \tilde{\Sigma}_{ij}$$

$$= e_j \frac{1}{\det \Sigma} \tilde{\Sigma} e_i^T$$

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

Q3.

3. Likelihood: $L(X_n \mid \mu, \Sigma) = \prod_{n=1}^{N} f_{x_n}(X_n ; \mu, \Sigma)$

→ log-likelihood:

$$l(X_n \mid \mu, \Sigma) = \log \prod_{n=1}^{N} f_{x_n}(X_n ; \mu, \Sigma)$$

$$\Rightarrow l(X_n \mid \mu, \Sigma) = \log \prod_{n=1}^{N} \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(X_n-\mu)^T \Sigma^{-1}(X_n-\mu)\right)$$

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

$$= -\frac{ND}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(X_n-\mu)^T\Sigma^{-1}(X_n-\mu)$$

→ since Σ is symmetric and independent on $(X_n-\mu)$

$$\Rightarrow \frac{\partial}{\partial \mu} l(X_n \mid \mu, \Sigma) = \sum_{n=1}^{N}\Sigma^{-1}(X_n-\mu) = 0.$$

→ since Σ is positive definite

$$\Rightarrow 0 = N\mu - \sum_{n=1}^{N} x_n$$

$$\Rightarrow \hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\Rightarrow \hat{\mu}_k = \frac{1}{N}\sum_{n=1}^{N} t_{nk}\, x_n$$

$$\rightarrow l(x_n \mid \mu, \Sigma) = \cdots - \frac{D}{2}\log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$$

$$= \cdots + \frac{D}{2}\log|\Sigma^{-1}| - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)^T\Sigma^{-1}(x_n-\mu)$$

$$\Rightarrow \frac{\partial}{\partial \Sigma^{-1}} l(x_n \mid \mu, \Sigma) = \frac{D}{2}\Sigma - \frac{1}{2}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T = 0.$$

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

$$\Rightarrow \hat{\Sigma} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)(x_n-\mu)^T = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk}(x_n-\mu_k)(x_n-\mu_k)^T$$

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

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