學號：T08902109　系級：　資工三　姓名：賈成銪

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

Kaggle 分數：

Generative Model :

Private Score : 0.84424　Public Score : 0.84324

Logistic Regression:

Private Score : 0.84842　Public Score : 0.85380

可以看出 logistic regression 的準確率要比 generative model 的準確率好。這與課堂上老師講的一般情況是一致的。

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

為做標準化處理時：

generative model：0.83970　logistic regression：0.78636

處理後：

generative model：0.84424　logistic regression：0.85380

可以看出特徵標準化對結果的影響很大，尤其是在 logistic regression，結果得到了很好的改進。

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

Best_model 我使用了 sklearn 的 SVM 模型。

通過調整多個參數觀察結果，最終得到的結果為在 C = 1.0 kernel = 'rbf'時最好。

結果：準確率 0.85921（public）　0.85480（private）

4. (3%) Refer to math problem

Next page:

1. denote the paraments of the model as θ

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

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

to maximize the log likelihood subject to the constraint $\sum_{k=1}^{k} \pi_k = 1$

we can write the Lagrangian fuction:

$$L(\pi, \lambda) = \sum_{n=1}^{N} \sum_{k=1}^{k} y_{n,k} [\ln P(x_n | C_k) + \ln \pi_k] + \lambda (\sum_{k=1}^{k} \pi_k - 1)$$

make:

$$\frac{d L(\pi, \lambda)}{d \pi_k} = 0 \qquad \frac{d(\pi, \lambda)}{d \lambda} = 0$$

$$= \frac{\sum_{n=1}^{N} y_{n,k}}{\pi_k} + \lambda \qquad = \sum_{k=1}^{k} \pi_k - 1$$

we have:

$$\sum_{n=1}^{N} y_{n,k} = -\lambda \pi_k \Rightarrow \pi_k = -\frac{\sum_{n=1}^{N} y_{n,k}}{\lambda} = -\frac{N_k}{\lambda}$$

$$\sum_{k=1}^{k} \pi_k = 1$$

$$\sum_{k=1}^{k} \pi_k = \sum_{k=1}^{k} -\frac{N_k}{\lambda} = 1 \Rightarrow \lambda = -N$$

Thus, we have that the maximum likelihood estimates of the prior probabilities are: $\pi_k = \frac{N_k}{N}$

2. $\frac{\partial \log|\Sigma|}{\partial \sigma_{ij}} = \frac{\partial|\Sigma|}{|\Sigma| \, \partial \sigma_{ij}}$

$|\Sigma| = \sum_j (-1)^{i+j} \sigma_{ij} M_{ij}$

the derivative of $|A|$ respect a_{ij} is $(-1)^{i+j} M_{ij}$

thus $\frac{\partial|\Sigma|}{|\Sigma| \, \partial \sigma_{ij}} = \frac{(-1)^{i+j} M_{ij}}{|\Sigma|} = \frac{e_j \tilde{\Sigma} e_i^T}{|\Sigma|} = e_j \Sigma^{-1} e_i^T$

3.

$P(x|C_k) = N(x|\mu_k, \Sigma)$

$L(\theta) = \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x^{(i)}-\mu_k)^T \Sigma^{-1} (x^{(i)}-\mu_k)\right\}$

$l(\theta) = \ln L(\theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (x^{(i)}-\mu_k)^T \Sigma^{-1} (x^{(i)}-\mu_k))$

$$l(\theta) = \sum_{k=1}^{K} -\frac{nN}{2} n(2\pi) - \frac{N}{2} \ln|\Sigma| - \frac{1}{2} \sum_{i=1}^{N} (x^{(i)} - \mu_k)^T \Sigma^{-1} (x^{(i)} - \mu_k)$$

① $$\frac{d(l(\mu, \Sigma | x^{(i)}))}{d\mu_k} = \sum_{i=1}^{N} \Sigma^{-1} (\mu_k - x^{(i)}) = 0$$

Σ is positive definite

thus: $N\mu_k = \sum_{i=1}^{N} x^{(i)} \Rightarrow \mu_k = \frac{\sum_{i=1}^{N} x^{(i)}}{N}$

thus: $$u_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} x_n$$

②
$$\frac{\partial(\mu, \Sigma | x^{(i)})}{\partial \Sigma^{-1}} = \sum_{k=1}^{K} \frac{N}{2}\Sigma - \frac{1}{2} \sum_{i=1}^{N} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T = 0$$

$$N \times N_k \Sigma = \sum_{i=1}^{N} \sum_{k=1}^{K} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T$$

thus $$N\Sigma = \sum_{k=1}^{K} \sum_{i=1}^{N} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T$$

$$= \sum_{k=1}^{K} N_k S_k$$

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