學號：R08246009　系級： 應數所 碩二　姓名：許哲維

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

logistic regression較佳(public, private)=(0.85651, 0.84756) ; 用generative model的表現只有 (public, private) = (0.8505, 0.83994), 差一點。

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

我們個別考慮generative model以及logistic regression對於做不做特徵化的差異，其他參數我們都選擇相同。首先是logistic regression (public, private) = (0.79361, 0.79326); 再來是generative model (public, private)=(0.85098, 0.83982)。可以發現做了normalization的參數對於logistic model的提昇是相當大的。

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

我使用gradient boosting搭配CV去做參數的選擇，最後選擇當max_depth=7, n_estimators=95 可以達到cv最佳的classification rate。

4. (3%) Refer to math problem

1.

$$P(X_i, C_k) = P(C_k) P(X_i \mid C_k)$$
$$= \pi_k P(X_i \mid C_k).$$

$$L(\pi_1, .., \pi_K \mid X_1, .., X_N)$$
$$= \prod_{n=1}^{N} [\pi_1 P(X_n \mid C_1)]^{t_{n1}} [\pi_2 P(X_n \mid C_2)]^{t_{n2}} \cdots [(1-\pi_1-\cdots-\pi_{K-1}) P(X_n \mid C_K)]^{1-\sum_{k=1}^{K-1} t_{nk}}$$

$$l(\pi) = \log L(\pi \mid X)$$

$$= \sum_{n=1}^{N} \left[\sum_{k=1}^{K-1} t_{nk} \log \pi_k + (1-\sum_{k=1}^{K-1} t_{nk}) \log(1-\pi_1-\cdots-\pi_{K-1}) \right]$$

$$+ (\quad \text{不重要} \quad)$$

$$\frac{\partial l(\pi)}{\partial \pi_k} = \sum_{n=1}^{N} \frac{t_{nk}}{\pi_k} - \frac{1-\sum_{k=1}^{K-1} t_{nk}}{1-\sum_{k=1}^{K-1} \pi_k} = 0.$$

$$\Rightarrow \frac{N_k}{\pi_k} = \frac{N-(N_1+\cdots+N_{K-1})}{1-\sum_{k=1}^{K-1} \pi_k} = 0 \quad \forall k=1, \dots, K-1$$

$$(N = N_1 + \cdots + N_K)$$

$$\Rightarrow \pi_k = \frac{N_k}{N} \quad \forall k=1, .., K-1$$

$$\left(\pi_K = 1 - \frac{\sum_{j=1}^{K-1} N_j}{N}\right)$$ #.

No.
Date

2.

Consider $\det\Sigma = \sum_{k=1}^{m} \sigma_{ik} C_{ik}$. (C_{ij} : is the cofactor matrix of Σ : $C_{ij} = (-1)^{i+j} M_{ij}$)

(M_{ij} : is the determinant of $(m-1)\times(m-1)$ matrix, which removes i-row, j-col.)

$$\frac{\partial \det\Sigma}{\partial \sigma_{ij}} = \sum_{k=1}^{m} \frac{\partial \sigma_{ik}}{\partial \sigma_{ij}} C_{ik} + \sigma_{ik} \underbrace{\frac{\partial C_{ik}}{\partial \sigma_{ij}}}_{= 0}$$

$$= C_{ij}$$

$$\Rightarrow \frac{\partial \log(\det\Sigma)}{\partial \sigma_{ij}} = \frac{1}{\det(\Sigma)} C_{ij} = (\Sigma^{-1})^T_{ij} = e_j \Sigma^{-1} e_i^T \quad \#$$

$$\left(\text{since } (\Sigma^{-1})^T_{ij} = \frac{1}{\det\Sigma} C_{ij} \right)$$

3.

$$p(X_i \mid C_k) = N(X_i \mid \mu_k, \Sigma)$$

Then the log likelihood function in Question 1. is given by

$$l(\mu, \Sigma) = \sum_{n=1}^{N} t_{n1} \log N(X_n \mid \mu_1, \Sigma) + \cdots + \left(1 - \sum_{i=1}^{k-1} t_{ni}\right) \log N(X_n \mid \mu_k, \Sigma)$$

$$\frac{\partial l(\mu, \Sigma)}{\partial \mu_1} = \sum_{n=1}^{N} t_{n1} \Sigma^{-1} (X_n - \mu_1) = 0$$

$$\Rightarrow \sum_{n=1}^{N} t_{n1} X_n = \mu_1 \sum_{n=1}^{N} t_{n1} = \mu_1 N_1$$

$$\Rightarrow \mu_1 = \frac{1}{N_1} \sum_{n=1}^{N} t_{n1} X_n$$

We can extend to $\underline{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} X_n}$ #

No.

Date :

$$\frac{\partial l(\mu, \Sigma)}{\partial \Sigma} = \sum_{n=1}^{N} t_{n1}\left[(\Sigma)^{T} + (x_n-\mu_1)(x_n-\mu_1)^{T}\right] + t_{n2}\left[(\Sigma)^{T} + (x_n-\mu_2)(x_n-\mu_2)^{T}\right]$$

$$+ \cdots + \left(1-\sum_{j=1}^{N-1} t_{nj}\right)\left[(\Sigma)^{T} + (x_n-\mu_k)(x_n-\mu_k)^{T}\right] = 0$$

$$\Rightarrow (N_1 + \cdots + N_k)\Sigma = \sum_{n=1}^{N}\left[t_{n1}(x_n-\mu_1)(x_n-\mu_1)^{T} + \cdots + \left(1-\sum_{j=1}^{N-1} t_{nj}\right)(x_n-\mu_k)(x_n-\mu_k)^{T}\right]$$

$$= N_1 S_1 + \cdots + N_k S_k$$

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

# #