ML HW2 Report

學號：B0705001 系級： 資管四 姓名：黃意芹

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

accuracy	generative model	logistic regression
Public score	0.80884	0.84805
Private score	0.80113	0.85270

在皆有normalization的情況下，Logistic regression的表現較佳。

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

generative model	Normalization	Without normalization
Public score	0.80884	0.84066
Private score	0.80113	0.83662

由此可見generative model做normalization反而表現變差。

logistic regression	Normalization	Without normalization
Public score	0.80884	0.76474
Private score	0.80113	0.76280

由此可見logistic regression做normalization表現大幅提升。

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sklearn的GradientBoostingClassifier實作，先把連續變數normalize，在特徵中我發現幾個continuous variable影響較明顯，所以把那幾項做平方、取exponential又加進特徵。之後放進GradientBoostingClassifier，經過多次validation測試，取了表現比較好的參數：learning_rate = 0.1, max_features = 0.9, max_depth = 3, n_estimators = 700作為模型。

4. (3%) Refer to math problem

No:
Date: /

1.

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

$$\frac{\partial L}{\partial \pi_k} = \frac{1}{\pi_k}\sum_{n=1}^{N} t_{nk} + \lambda = 0 \rightarrow \frac{1}{\pi_k} N_k = -\lambda \rightarrow \pi_k = -\frac{N_k}{\lambda} \text{ --- ①}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{k=1}^{K}\pi_k - 1 = 0 \rightarrow \sum_{k=1}^{K}\pi_k = 1 \text{ --- ②}$$

$$\sum_{k=1}^{K}\pi_k \overset{\text{由①}}{=} \sum_{k=1}^{K} -\frac{N_k}{\lambda} = -\frac{N}{\lambda} \overset{\text{由②}}{=} 1 \Rightarrow \lambda = -N \text{ 代入 ①}$$

$$\therefore \pi_k = \frac{N_k}{N} \text{ ※}$$

No:

Date: /

2\. 左式: $\dfrac{\partial \log|\Sigma|}{\partial \sigma_{ij}} = \dfrac{1}{|\Sigma|}\dfrac{\partial|\Sigma|}{\partial\sigma_{ij}}$

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

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

where M_{ij} is the determinant of the matrix obtained by removing ith row and jth column of Σ.

j th row of adj(Σ) ↑ 即 j th col of $\tilde{\Sigma}$

右式 $e_j\Sigma^{-1}e_i^T = e_j\dfrac{\mathrm{adj}(\Sigma)}{|\Sigma|}e_i^T = \dfrac{1}{|\Sigma|}(-1)^{i+j}M_j\, e_i^T$

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

$$\because \begin{array}{c} \begin{bmatrix} 0 & 0 & \boxed{1} & \cdots & 0 \end{bmatrix} \\ \quad\ \ j \end{array} \begin{bmatrix} (-1)^2 M_{11} & (-1)M_{21} & \cdots & M_{n1} \\ \vdots & \vdots & & \\ (-1)^{1+j}M_{1j} & (-1)^{2+j}M_{2j} & \cdots & M_{nj} \\ & & & \end{bmatrix} \begin{array}{c} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \boxed{1} \\ 0 \end{bmatrix} \end{array} \begin{array}{c} \\ \\ \\ i \\ \\ \end{array} = (-1)^{i+j}M_{ij}$$

where $\mathrm{adj}(\Sigma) = \tilde{\Sigma}^T = (-1)^{i+j}\times M_{ji} = \begin{bmatrix} (-1)^{1+1}M_{11} & (-1)^{12}M_{21} & \cdots \\ \vdots & & \vdots \\ \cdots & \cdots & (-1)^{n+n}M_{nn} \end{bmatrix}$

左式=右式 故得證 ✱

No:
Date: /

3. log Likelihood function with terms that depend on μ_k:

$$L=\sum_{n=1}^{N}\sum_{k=1}^{K}\ln N(x_n|\mu_k,\Sigma)=-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n-\mu_k)^T\Sigma^{-1}(x_n-\mu_k)+C$$

(C: constant)

$$\frac{\partial L}{\partial \mu_k}=0\rightarrow \frac{\partial \sum_{n=1}^{N}t_{nk}(x_n-\mu_k)^T\Sigma^{-1}(x-\mu_k)}{\partial \mu}=0$$

$\because \frac{\partial w^TAw}{\partial w}=2Aw$ if w does not depend on A and A is symmetric

$\rightarrow \sum_{n=1}^{N}t_{nk}\Sigma^{-1}(x-\mu_k)=0$ $\because \Sigma$ is positive definite

$\rightarrow N_k\mu_k-\sum_{n=1}^{N}t_{nk}x_n=0 \rightarrow \mu_k=\frac{1}{N_k}\sum_{n=1}^{N}t_{nk}x_n$ #

$$L=\sum_{n=1}^{N}\sum_{k=1}^{K}-\frac{1}{2}t_{nk}(x_n-\mu_k)^T\Sigma^{-1}(x_n-\mu_k)-\frac{N_k}{2}\ln|\Sigma|+C$$

$$=\frac{N_k}{2}\log|\Sigma^{-1}|-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}\,tr[(x_n-\mu_k)(x_n-\mu_k)^T\Sigma^{-1}]$$

$$\frac{\partial L}{\partial \Sigma^{-1}}=-\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n-\mu_k)(x_n-\mu_k)^T+\frac{N}{2}\Sigma+C \quad \because \Sigma^T=\Sigma$$

$$\rightarrow N\Sigma-\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n-\mu_k)(x_n-\mu_k)^T=0$$

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

$$\rightarrow \Sigma=\frac{N_k}{N}\cdot\frac{\sum_{n=1}^{N}\sum_{k=1}^{K}t_{nk}(x_n-\mu_k)(x_n-\mu_k)^T}{N_k}=\frac{N_k}{N}S_k$$ #