學號：B06507002 系級： 材料三 姓名：林柏勳

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

model\Accuracy	3-fold cross validation
gemerative model	0.80593
logistic regression	0.85387

logistic regression的分數較佳，原因在於generative model中，我們假設機率的分布是高斯，但真實世界可能不是高斯分布。相較之下，logistic regression 較沒有此問題，它單純用gradient descent的方法去尋找最佳解，所以表現較好。

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

以logistic regression 為例，我用 3-fold cross validation作為指標，比對有沒有加特徵標準化的結果如下:

特徵標準化	有加	沒有加
logistic regression	0.85387	0.83711

以cross-validation的分數看來，有加特徵標準化這個行為對於模型的預測準確率是有幫助的。

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

a. 做 feature engineering: 刪掉國家以及、fnlwgt這兩個因子，並以合併縮小feature的數目，e.g.

Married=Married-AF-spouse+Married-civ-spouse+Married-spouse-absent

UnMarried=Separated+Widowed+Divorced

SL-gov=Local-gov+State-gov

Self-emp=Self-emp-inc+Self-emp-not-inc

Unemployeed=Never-worked+Without-pay

b. 做 normalization

c. 手刻 gradient boost decision tree (GBDT) 後以cross-validation尋找最佳的hyperparameters , 得到learning rate =0.1, maximum depth=4, number of trees=250

準確率如下:

train	public	private
0.88197	0.87739	0.87176

參考資料: https://arxiv.org/ftp/arxiv/papers/1810/1810.10076.pdf

4. (3%) Refer to math problem

HW 2

1. generative classification model for K classes

$p(C_k) = \pi_k \quad , \quad p(X \mid C_k)$

train $\{X_n, t_n\}$ $\quad t_n$: binary target

$P(X_n, C_k) = \underbrace{P(C_k)}_{\pi_k} P(X_n \mid C_k) \approx f_{\mu_k, \Sigma}(X_n)$

likelihood function

$$L(t \mid \mu_1, \dots \mu_K, \Sigma, \pi_1, \dots \pi_{K-1}) = \prod_{n=1}^{N} \prod_{j=1}^{K} [f_{\mu_j, \Sigma}(X_n) \pi_j]^{t_{nj}}$$

$t_{nj} \in (N, K)$

$$\ln L(t \mid \mu_1, \dots \mu_K, \Sigma, \pi_1, \dots \pi_{K-1}) = \sum_{n=1}^{N} \sum_{j=1}^{K} t_{nj} \ln f_{\mu_j, \Sigma}(X_n) + t_{nj} \ln \pi_j$$

where $\pi_K = 1 - \sum_{i=1}^{K-1} \pi_i$

$$\frac{\partial \ln L}{\partial \pi_k} = \sum_{n=1}^{N} \frac{\partial}{\partial \pi_k} \{ t_{nk} \ln f_{\mu_k, \Sigma}(X_n) + t_{nk} \ln \pi_k + t_{nK} \ln f_{\mu_K, \Sigma}(X_n) + t_{nK} \ln (1 - \textstyle\sum_{j=1}^{K-1} \pi_j) \}$$

$$= \sum_{n=1}^{N} \frac{t_{nk}}{\pi_k} + \frac{t_{nK}}{1 - \sum_{j=1}^{K-1} \pi_j} (-1) = 0$$

$$\Rightarrow \frac{\sum_{n=1}^{N} t_{nk}}{\pi_k} = \frac{\sum_{n=1}^{N} t_{nK}}{1 - \sum_{j=1}^{K-1} \pi_j} \Rightarrow \frac{N_k}{\pi_k} = \frac{N_K}{1 - \sum_{j=1}^{K-1} \pi_j}$$

As a result,

$$\frac{N_1}{\pi_1} = \frac{N_2}{\pi_2} = \dots = \frac{N_k}{\pi_k} = \dots = \frac{N_K}{1 - \sum_{j=1}^{K-1} \pi_j}$$

let $\frac{N_1}{\pi_1} = a \Rightarrow N = N_1 + \dots N_K$

$$= a(\cancel{\pi_1 + \pi_2 + \dots \pi_{K-1}} + 1 - \cancel{\textstyle\sum_{j=1}^{K-1} \pi_j}) = a$$

$\frac{N_k}{\pi_k} = a = N$

$\Rightarrow \pi_k = \frac{N_k}{N}$ ※

GEE-JUMP

2.

Show that $\dfrac{\partial \log |\Sigma|}{\partial \sigma_{ij}} = e_j \Sigma^{-1} e_i^T$

$\Sigma \in \mathbb{R}^{m \times m}$, $e_j \in$ row vector

ex. $e_3 = [0 \ 0 \ 1 \ 0 \ \cdots \ 0]$

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

$$= \frac{1}{|\Sigma|} \frac{\partial |\Sigma|}{\partial \sigma_{ij}} = \frac{C_{ij}}{|\Sigma|} = \frac{\mathrm{adj}(A)_{ji}}{|\Sigma|} = (\Sigma^{-1})_{ji}$$

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

3. $p(X \mid C_k) = N(X \mid \mu_k, \Sigma)$

Show $\mu_k = \dfrac{1}{N_k} \sum_{n=1}^{N} t_{nk} X_n$, $\Sigma = \sum_{k=1}^{K} \dfrac{N_k}{N} S_k$

where $S_k = \dfrac{1}{N_k} \sum_{n=1}^{N} t_{nk} (X_n - \mu_k)(X_n - \mu_k)^T$

$$L(t \mid \mu_1, \dots \mu_K, \Sigma, \pi_1, \dots \pi_{K-1})$$

$$= \prod_{n=1}^{N} \prod_{j=1}^{K} \{ f_{\mu_j, \Sigma}(X_n) \pi_j \}^{t_{nj}}$$

$$\ln L = \sum_{n=1}^{N} \sum_{j=1}^{K} \{ t_{nj} \ln f_{\mu_j, \Sigma}(X_n) + t_{nj} \ln \pi_j \}$$

$$\frac{\partial \ln L}{\partial \mu_k} = \sum_{n=1}^{N} \frac{\partial}{\partial \mu_k} \{ t_{nk} \ln f_{\mu_k, \Sigma}(X_n) + t_{nk} \ln \pi_k \}$$

$$\frac{\partial \ln L}{\partial \mu_k} = \sum_{n=1}^{N} t_{nk} \frac{1}{f_{\mu_k,\Sigma}(x_n)} \frac{\partial f_{\mu_k,\Sigma}(x_n)}{\partial \mu_k} = 0$$

$$= \sum_{n=1}^{N} t_{nk} \frac{\partial}{\partial \mu_k} \left\{ -\frac{1}{2} (x_n - \mu_k)^T (\Sigma)^{-1} (x_n - \mu_k) \right]$$

$$= \sum_{n=1}^{N} t_{nk} \left(-\frac{1}{2}\right) \left\{ \frac{\partial (x_n - \mu_k)}{\partial \mu_k} (\Sigma)^{-1} (x_n - \mu_k) + \frac{\partial (x_n - \mu_k)}{\partial \mu_k} \left\{(\Sigma)^{-1}\right]^T (x_n - \mu_k) \right]$$

where $(\Sigma^{-1})^T = \Sigma^{-1}$ since Σ is symmetric

$$= \sum_{n=1}^{N} t_{nk} (x_n - \mu_k) = 0$$

$$\Rightarrow \sum_{n=1}^{N} t_{nk} x_n = N_k \mu_k \Rightarrow \mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} x_n$$

※

$$\frac{\partial \ln L}{\partial \Sigma} = \sum_{n=1}^{N} \sum_{j=1}^{K} \left\{ t_{nj} \frac{\partial \ln f_{\mu_j,\Sigma}(x_n)}{\partial \Sigma} \right]$$

① $$f_{\mu_j,\Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T (\Sigma)^{-1} (x-\mu)\right)$$

$$\frac{\partial f_{\mu,\Sigma}(x)}{\partial \Sigma} = \frac{1}{(2\pi)^{D/2}} \left(-\frac{1}{2}\right) |\Sigma|^{\cancel{-3/2}\,-\frac{1}{2}} \cancel{|\Sigma|} (\Sigma^{-1})^T \cdot \exp\left(-\frac{1}{2}(x-\mu)^T (\Sigma)^{-1} (x-\mu)\right) +$$

$$\boxed{\frac{\partial |A|}{\partial A} = |A| (A^{-1})^T}$$

$$\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T (\Sigma)^{-1} (x-\mu)\right) \cdot \left(-\frac{1}{2}\right)(-1)(\Sigma^{-1})^T (x-\mu)(x-\mu)^T (\Sigma^{-1})^T$$

$$= f_{\mu_j,\Sigma}(x) \left\{ -\frac{1}{2}(\Sigma^{-1})^T + \frac{1}{2}(\Sigma^{-1})^T (x-\mu)(x-\mu)^T (\Sigma^{-1})^T \right]$$

③ date . . No.

$$\frac{\partial \ln L}{\partial \Sigma} = \frac{1}{2}\sum_{n=1}^{N}\sum_{j=1}^{K} t_{nj} \left[(\Sigma^{-1})^{T} (x_n - \mu_j)(x_n - \mu_j)^{T} (\Sigma^{-1})^{T} - (\Sigma^{-1})^{T} \right]$$

for each class k, it has terms of

$$\frac{1}{2}\sum_{n=1}^{N} t_{nk} \left[(\Sigma^{-1})(x_n - \mu_j)(x_n - \mu_j)^{T}(\Sigma^{-1}) - (\Sigma^{-1}) \right]$$

$$= \frac{1}{2}\Sigma^{-1}\left[\sum_{n=1}^{N} t_{nk}(x_n - \mu_j)(x_n - \mu_j)^{T}\right](\Sigma^{-1}) - \frac{1}{2}N_k\Sigma^{-1}$$

$$= \frac{1}{2}\Sigma^{-1}N_k S_k \Sigma^{-1} - \frac{1}{2}N_k\Sigma^{-1}$$

for all class

$$\cancel{\frac{1}{2}}\sum_{k=1}^{K}\left[\Sigma^{-1}N_k S_k\Sigma^{-1} - N_k\Sigma^{-1}\right] = \cancel{0}$$

$$\Rightarrow \Sigma^{-1}\left(\sum_{k=1}^{K} N_k S_k\right)\Sigma^{-1} - N\Sigma^{-1} = 0$$

$$\Rightarrow \left[\Sigma^{-1}\left(\sum_{k=1}^{N} N_k S_k\right) - N\right]\Sigma^{-1} = 0$$

$$\underbrace{}_{=0}$$

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

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