ML HW2 Report

學號：R08942025　系級：電信碩一　姓名：徐瑋辰

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

Model	generative model		logistic regression	
Accuracy	public	private	public	private
	0.84422	0.84412	0.85442	0.84903

由上表數據，logistic regression(有做 feature normalization)的結果較佳

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

作法：在 age, fnlwgt, capital-gain, capital-loss, hours-per-week 這五項連續的參數做標準化

影響：這五項特徵的"距離基準"(標準差)轉為一致，可視為其對最終結果的影響程度相同，在做 gradient descent 的時候，可以使參數收斂較快

結果：

Model	with feature normalization		without feature normalization	
Accuracy	public	private	public	private
	0.85442	0.84903	0.79606	0.79081

由上表數據，有做 feature normalization 的 logistic regression 結果較佳

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

本作業的 best model 採用 sklearn 函式庫中的 GradientBoostingClassifier。它是運用 Gradient boosting Decision Tree(GBDT)這個演算法來做回歸。其原理是在訓練過程中，根據每次預測錯誤的資料去做調整，使 Loss function 在訓練過程中逐漸降低。最終訓練的模型則是將前述的所有模型間加權組合而成。其優勢是相較於傳統的 linear regression，可以比較快收斂。

另外，在去除 fnlwgt(標準差過大)及國籍(分類太多)這兩項特徵後，加上特徵標準化，可以得到更好的結果。

Model	GBDT(without fnlwgt)		GBDT(without fnlwgt, nationality)	
Accuracy	public	public	public	private
	0.86242	0.86781	0.86904	0.86230

4. (3%) Refer to math problem

1. $P(x,t) = P(x|t)P(t) = \prod_{k=1}^{K} (P(x|C_k)\cdot \pi_k)^{t_k}$

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

$$\xrightarrow{\log} l(\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} t_{n,k}[\log(P(x_n|C_k)) + \log \pi_k]$$

by Lagrange Multipliers

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

$$\frac{\partial}{\partial \pi_k} L(\pi,\lambda) = \frac{1}{\pi_k}\sum_{n=1}^{N} t_{n,k} + \lambda = 0 \Rightarrow \pi_k = -\frac{1}{\lambda}\sum_{n=1}^{N} t_{n,k} = \frac{-N_k}{\lambda}$$

$$\frac{\partial}{\partial \lambda} L(\pi,\lambda) = \sum_{k=1}^{K} \pi_k - 1 = 0 \Rightarrow \sum_{k=1}^{K} \pi_k = 1$$

$$\sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K}\left(-\frac{N_k}{\lambda}\right) = -\frac{N}{\lambda} = 1 \Rightarrow \lambda = -N$$

$$\Rightarrow \pi_k = \frac{-N_k}{\lambda} = \frac{-N_k}{-N} = \underline{\frac{N_k}{N}} \#$$

2. $\dfrac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = e_j \Sigma^{-1} e_i^T$

$$= \frac{1}{\det(\Sigma)} \boxed{\frac{\partial(\det \Sigma)}{\partial \sigma_{ij}}} \qquad = \underline{e_j \Sigma^{-1} e_i^T} \#$$

$$= \frac{1}{\det(\Sigma)} \frac{\partial \Sigma_j (\sigma_{ij} \Sigma_{ij})}{\partial \sigma_{ij}}$$

$$= \frac{1}{\det(\Sigma)} \cdot \sigma_{ij}$$

$$= \frac{1}{\det(\Sigma)} \cdot (\det(\Sigma)(\Sigma^{-1})^T)_{ij}$$

$$= \frac{1}{\cancel{\det(\Sigma)}} \cancel{\det(\Sigma)} \cdot (\Sigma^{-1})^T_{ij}$$

3 $p(x|C_k) = N(x|\mu_k, \Sigma) \rightarrow$

(1) $$l(\mu_k, \Sigma | x_n) = \log \prod_{n=1}^{N} \frac{t_{nk}}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k)\right)$$

$$= \sum_{n=1}^{N} t_{nk} \left(-\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k)\right)$$

$$\rightarrow l(\mu_k, \Sigma) = \frac{N \cdot k \cdot t_{nk}}{-2} \log(2\pi) - \frac{N t_{nk}}{2} \log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N} (x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k) t_{nk}$$

$$\frac{\partial l(\mu_k, \Sigma)}{\partial \mu_k} = \sum_{n=1}^{N} \Sigma^{-1} (\mu_k - x_n) t_{n,k} = 0$$

$$\rightarrow 0 = \mu_k \underbrace{\boxed{\sum_{n=1}^{N} t_{n,k}}}_{} ^{= N_k} - \sum_{n=1}^{N} x_n t_{n,k}$$

$$\rightarrow \mu_k = \frac{1}{N_k} \sum_{n=1}^{N} x_n t_{n,k}$$

3 (2) 承上

$$l(\mu_k, \Sigma | x_n) = \sum_{n=1}^{N} t_{n,k} \left(-\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k)\right)$$

$$= \frac{-Nk}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{n=1}^{N} t_{n,k}(x_n - \mu_k)^T \Sigma^{-1} (x_n - \mu_k)$$

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

$$\frac{\partial l(\mu_k, \Sigma | x_n)}{\partial \Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}\sum_{n=1}^{N} t_{n,k}(x_n - \mu_k)(x_n - \mu_k)^T = 0$$

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

$$\Sigma = \frac{1}{N}\sum_{n=1}^{N} t_{n,k}(x_n - \mu_k)(x_n - \mu_k)^T$$

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

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