NO.
Date

學號 R08921A01 系級 電機碩一 姓名 陳OO

1. logistic regression

2. 標準化後準確率提高。

3. (1) 捨棄 fnlwgt 和 education_num 這 2 個 feature
(2) 把 age 除以 2 取 int、hours_per_week 除以 3 取 int 再做 one hot encoding
(3) capital_gain、capital_loss 開根号
(4) 所有 feature 都標準化
(5) 產生類別變項的交互項
(6) 所有 feature 只取相關係數 >0.05 or <-0.05
(7) feature 中若有 2 個相關係數 >0.95 只留下一個
(8) 使用 sklearn.ensemble.GradientBoostingClassifier
(9) public : 0.86941

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

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

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

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

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

2. Show that $\frac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = e_j \Sigma^{-1} e_i^T$, $\Sigma \in \mathbb{R}_{m \times m}$

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

$$\frac{\partial \log|\Sigma|}{\partial \sigma_{ij}} = \frac{\partial \log\left(\sum_{j=1}^{m} (-1)^{i+j} \sigma_{ij} M_{ij}\right)}{\partial \sigma_{ij}} = \frac{1}{|\Sigma|} \cdot (-1)^{i+j} M_{ij}$$

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

$$\frac{\partial \log(\det \Sigma)}{\partial \sigma_{ij}} = \frac{1}{|\Sigma|} (-1)^{i+j} M_{i+j} = \frac{1}{|\Sigma|} \hat{\Sigma}_{ij} = \Sigma^{-1}_{ij} = e_j \Sigma^{-1} e_i^T$$

3. $$L(\mu_k, \Sigma | x_1 x_2 \cdots x_N) = \log \prod_{i=1}^{N} \frac{t_{ik}}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma^{-1} (x_i - \mu_k)\right)$$

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

$$\frac{\partial L(\mu_k \Sigma | x_1 x_2 \cdots x_N)}{\partial \mu_k} = \sum_{i=1}^{N} t_{ik} \Sigma^{-1} (\mu_k - x_i) = 0$$

$$0 = N_k \mu_k - \sum_{i=1}^{N} t_{ik} x_i \Rightarrow \boxed{\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} t_{nk} x_n}$$ ※

$$\frac{\partial L(\mu, \Sigma | x_1 x_2 \cdots x_N)}{1} = \sum_{n=1}^{N} \frac{-k}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2} tr\left[(x_n - \mu)(x_n - \mu)^T \Sigma^{-1}\right]$$

$$= \left(\sum_{n=1}^{N} \frac{-k}{2} \log(2\pi)\right) - \frac{1}{2}\left(N \log|\Sigma| + \sum_{k=1}^{K} tr[S_k N_k \Sigma^{-1}]\right)$$

$$\frac{\partial L(\mu, \Sigma | x_1 x_2 \cdots x_N)}{\partial \Sigma} = \frac{-1}{2}\left(N\Sigma^{-1} + \sum_{k=1}^{K}\left(-\Sigma^{-1} S_k N_k \Sigma^{-1}\right)\right) = 0$$

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