HW2 Report

學號：B05902017 系級： 資工四 姓名：王盛立

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

我的logistic regression的準確率比generative model的準確率高。在logistic regression中，他的sigmoid funcion讓比較極端的資料能控制在一定的範圍，因此相較於generative的表現要好一些。

2. 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

我有測驗過不做任何特徵標準化，和將特徵以$x - mean(x)/std(x)$的方式還有 $x - min(x)/max(x) - min(x)$兩種方式做特徵標準化。如果不做特徵標準化的話，準確率會下降滿多的，做之後會有顯著的改善。其中$x - mean(x)/std(x)$會比較好，可能是因為 $x - min(x)/max(x) - min(x)$會把特徵壓到0到1間，反而影響到特徵的影響力。

3. 請說明你實作的best model，其訓練方式和準確率為何?

我將幾個連續數字的feature，例如age、capital gain等等不是只有0跟1的feature去取2到5次方，還有sin,cos,tan等等。 一樣用logistic regression去跑，準確率從0.855上升到了0.865。另外我也有試著用neural network的方式去訓練，但最後結果似乎差不多我就沒有放上來了。

4. Refer to math problem

1

令 x_n屬於class C_{x_n}

Its likelihood function is $P(x_1, x_2, x_3, \cdots, x_N)$ = $\pi_{n=1}^{N} P(x_n)$ = $\pi_{n=1}^{N} P(C_{x_n})P(x_n|Cx_n)$

Its log likelyhood fuction is $logP(x_1, x_2, x_3, \cdots, x_N)$ =

$\sum_{n=1}^{N} logP(C_{x_n}) + \sum_{n=1}^{N} logP(x_n|Cx_n)$ = $\sum_{k=1}^{K} N_k logP(C_k) + \sum_{n=1}^{N} logP(x_n|Cx_n)$ =

$\sum_{k=1}^{K} N_k log\pi_k + \sum_{n=1}^{N} logP(x_n|Cx_n)$

最大化log likelyhood function就可得likelihood function的最大值，另外我們已知 $\sum_{k=1}^{K} \pi_k = 1$

令 $f = logP(x_1, x_2, \cdots, x_N)$ and $g = \sum_{k=1}^{K} \pi_k = 1$

$\frac{\partial}{\partial \pi_i} f = \frac{\partial}{\partial \pi_i} (\sum_{k=1}^{K} N_k log \pi_k + \sum_{k=1}^{N} log P(x_n | C_{x_n})) = \frac{\partial}{\partial \pi_i} \sum_{k=1}^{K} N_k log \pi_k + 0$

$= \frac{\partial}{\partial \pi_i} N_i log \pi_i = \frac{N_i}{\pi_i}$

$\Rightarrow \nabla f$ = $\begin{bmatrix} \frac{\partial}{\partial \pi_1} f \\ \frac{\partial}{\partial \pi_2} f \\ \vdots \\ \frac{\partial}{\partial \pi_k} f \end{bmatrix}$ = $\begin{bmatrix} \frac{N_1}{\pi_1} \\ \frac{N_2}{\pi_2} \\ \vdots \\ \frac{N_k}{\pi_k} \end{bmatrix}$

$\frac{\partial}{\partial \pi_i} g$ = $\frac{\partial}{\partial \pi_i} \sum_{k=1}^{K} \pi_k$ = $\frac{\partial}{\partial \pi_i} \pi_i$ = 1

$\Rightarrow \nabla g$ = $\begin{bmatrix} \frac{\partial}{\partial \pi_1} g \\ \frac{\partial}{\partial \pi_2} g \\ \vdots \\ \frac{\partial}{\partial \pi_k} g \end{bmatrix}$ = $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

若假設∇f = $\lambda \nabla g$，得 $\begin{bmatrix} \frac{N_1}{\pi_1} \\ \frac{N_2}{\pi_2} \\ \vdots \\ \frac{N_k}{\pi_k} \end{bmatrix}$ = $\lambda \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

$\Rightarrow \pi_i$ = $\frac{N_i}{\lambda}$ ，且$\sum_{k=1}^{K} \pi_k$ = $\sum_{k=1}^{K} \frac{N_k}{\lambda}$ = $\frac{N}{\lambda}$ = 1

可得λ = N，所以$\pi_i = \frac{N_i}{N}$

2

$\frac{\partial log(det\ \mathbf{\Sigma})}{\partial \sigma_{ij}}$ = $\frac{1}{det\ \mathbf{\Sigma}} \frac{\partial}{\partial \sigma_{ij}} (\sigma_{i1} c_{i1} + \sigma_{i2} c_{i2} + \sigma_{i3} c_{i3} + \cdots + \sigma_{ij} c_{ij} + \sigma_{im} c_{im})$

= $\frac{1}{det\ \mathbf{\Sigma}} C_{ij}$ = $\frac{1}{det\ \mathbf{\Sigma}} adj(\Sigma)_{ji}$ = $(\Sigma^{-1})_{ji} = e_j \Sigma^{-1} e_i^T$

3

由第一題知log lifelihood function為$log P(x_1, x_2, \cdots, x_N)$ =

$\sum_{k=1}^{K} N_k log \pi_k + \sum_{n=1}^{N} log P(x_n | C x_n)$ = $\sum_{k=1}^{K} N_k log \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} lnk log P(x_n | C_k)$

= $\sum_{k=1}^{K} N_k log \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} lnk log N(x_n | \mu_k \Sigma)$

令a = $\sum_{k=1}^{K} N_k log \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} lnk log N(x_n | \mu_k \Sigma)$

$\frac{\partial a}{\partial \mu_i}$ = $\frac{\partial}{\partial \mu_i} \sum_{k=1}^{K} \sum_{n=1}^{N} lnk(-\frac{1}{2}(\mu_k - x_n)^T \Sigma^{-1} (\mu_k - x_n) - \frac{1}{2} logdet\Sigma - \frac{m}{2} log2\pi)$ =

$\frac{\partial}{\partial \mu_i} \sum_{n=1}^{N} lni(-\frac{1}{2}(\mu_i - x_n)^T \Sigma^{-1} (\mu_i - x_n) - \frac{1}{2} logdet\Sigma - \frac{m}{2} log2\pi)$ =

$\sum_{n=1}^{N} lni(-\frac{1}{2} 2(\mu_i - x_n)\Sigma^{-1}$ = $(\sum_{n=1}^{N} ln_i x_n^T - (\sum_{n=1}^{N} ln_i)\mu_i^T)\Sigma^{-1}$ =

$(\sum_{n=1}^{N} ln_i x_n^T - N_i \mu_i^T)\Sigma^{-1}$

⇒ $(\sum_{n=1}^{N} ln_i x_n^T - (\sum_{n=1}^{N} ln_i)\mu_i^T)\Sigma^{-1}$ = 0

⇒ μ_i = $\frac{1}{N_i} \sum_{n=1}^{N} ln_i X_n$

$\frac{\partial a}{\partial \Sigma^{-1}}$ = $\frac{\partial}{\partial \mu_i} \sum_{k=1}^{K} \sum_{n=1}^{N} lnk(-\frac{1}{2}(\mu_k - x_n)^T \Sigma^{-1} (\mu_k - x_n) - \frac{1}{2} logdet\Sigma - \frac{m}{2} log2\pi)$ =

$\sum_{k=1}^{K} \sum_{n=1}^{N} lnk(-\frac{1}{2}(\mu_k - x_n)(\mu_k - x_n)^T - \frac{1}{2}(-\Sigma))$ =

$\frac{1}{2} \sum_{k=1}^{K} \sum_{n=1}^{N} (ln_k \Sigma - ln_k (\mu_k - x_n)(\mu_k - x_n)^T)$ = $\frac{1}{2} \sum_{k=1}^{K} ((\sum_{n=1}^{N} ln_k)\Sigma - N_k S_k)$ =

$\frac{1}{2} \sum_{k=1}^{K} (N_k \Sigma - N_k S_k)$ = $\frac{1}{2}(N\Sigma - \sum_{k=1}^{K} N_k S_k)$

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