Homework 2 Report

學號：r07942086　系級：電信碩二　姓名: 顏宏宇

1. (1%) 請說明你實作之 logistic regression 以及 generative model 於此 task 的表現，並試著討論造成此差異及可能原因。

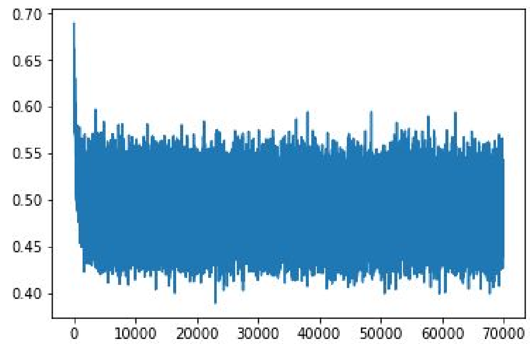
ans0.5122442370864515.csv a minute to go by r07942086_hyyen logistic regression	0.81660	0.81940
ansg.csv 30 minutes ago by r07942086_hyyen generative model	0.80760	0.80960

如上圖，我的generative model對於這個task比較不準確，可能是因為logistic regression model我有做比較多的改良(adam, data standardization等)，而generative model我只有用基本的機率去算，所以效果比較不好。

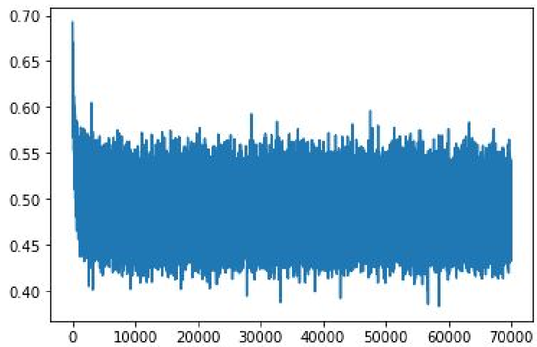
2. (1%) 請試著將 input feature 中的 gender, education, martial status 等改為 one-hot encoding 進行 training，並比較其與原本的差異及其可能原因。

兩者的訓練過程如下，基本上看起來差不多。

沒有one-hot encoding



有one-hot encoding



而最後的準確率，可以看到雖然public是沒有one-hot比較高，但private卻是有one-hot比較高，所以我想對我的模型而言，有沒有one-hot的效果其實差不多。
可能是因為原本的data就足以train到local minimum，就算改了one-hot還是會卡在差不多的local minimum，改模型設計比較可能有比較大的影響。

ans0.47696870168608513.csv 4 days ago by r07942086_hyyen with one hot	0.82000	0.81680
ans0.44829247485391494.csv 4 days ago by r07942086_hyyen no one hot	0.81620	0.81980

3. (1%)請試著討論哪些 input features 的影響較大、哪些input features 的影響較小（方法不限）。

ans0.43573680484651506.csv 2 days ago by r07942086_hyyen no PAY_AMT1 - PAY_AMT6	0.81760	0.81900
ans0.4781551698106972.csv 2 days ago by r07942086_hyyen no BILL_AMT1 - BILL_AMT6	0.81980	0.81860
ans0.5059310074026229.csv 3 days ago by r07942086_hyyen no PAY0 - PAY6	0.78140	0.78160
ans0.49396942775800873.csv 3 days ago by r07942086_hyyen no AGE	0.81700	0.82040
ans0.5060405672602863.csv 3 days ago by r07942086_hyyen no MARRIAGE	0.81600	0.81860
ans0.48269579698883847.csv 3 days ago by r07942086_hyyen no EDUCATION	0.81580	0.82000
ans0.47791390190311295.csv 3 days ago by r07942086_hyyen no SEX	0.81720	0.81920
ans0.48222005536708906.csv 4 days ago by r07942086_hyyen no LIMIT_BAL	0.81620	0.81740

我的作法是把各種input feature從training data拿掉，看拿掉哪一種feature會讓模型效果變差最多，就是對模型影響最大的feature。

可以看到PAY0 ~PAY6的feature平均分數最低，對模型影響最大。
BILL_AMT1~BILL_AMT6的feature平均分數最高，對模型影響最小。

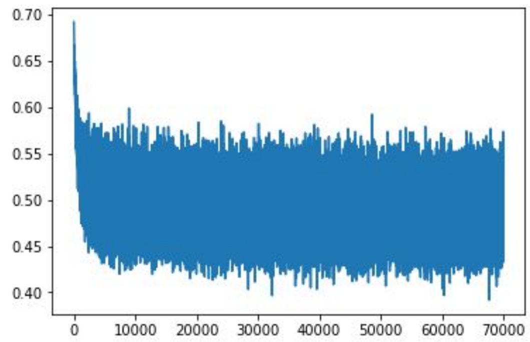
4. (1%)請實作特徵標準化 (feature normalization),並討論其對於模型準確率的影響與可能原因。

兩者的準確率如下圖，可以看到有feature normalization的效果明顯好很多。

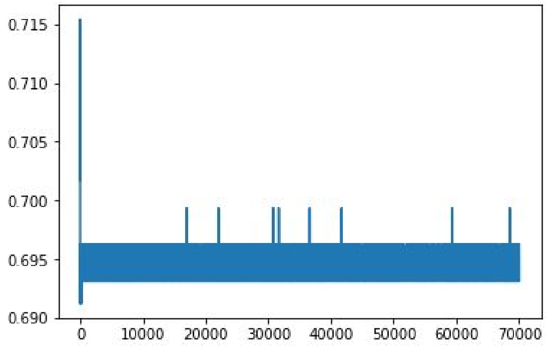
ans0.6962475522356161.csv 4 days ago by r07942086_hyyen Without feature normalization	0.21880	0.21940
ans0.47872810863806065.csv 4 days ago by r07942086_hyyen With feature normalization	0.72440	0.74160

而兩者的訓練過程(loss)如下，可以看出對於沒有feature normalization的版本在訓練過程中 loss很容以卡在差不多的值，可能是陷入local minimum或是平坦的地方，而且難以走出來。
而有feature normalization 之後由於feature之間的scale差異比較小，比較容易train出比較好的結果。

有 feature normalization



沒有 feature normalization



5-6題 collaborator : d07946003 王嘉澤

5.

5.

Consider $I=\int_{-\infty}^{\infty} e^{-x^2}\,dx$, $I>0$

$\Rightarrow I^2=\left(\int_{-\infty}^{\infty} e^{-x^2}\,dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\,dy\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\,dx\,dy$

let $x=r\cos\theta$, $y=r\sin\theta$

$r\in[0,\infty]$, $\theta\in[0,2\pi]$

$\Rightarrow I^2=\int_0^{2\pi}\int_0^{\infty} e^{-r^2}\,r\,dr\,d\theta = 2\pi\cdot\frac{1}{2}=\pi$

$\Rightarrow I=\sqrt{\pi}$

Thus. $f(x)=\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2}dx = \frac{\sqrt{2}\,\sigma}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-t^2}\,dt=\frac{\sqrt{\pi}}{\sqrt{\pi}}=1$$ #

6.

6.

(a) $$\frac{\partial E}{\partial z_k}=\frac{\partial E}{\partial y_k}\cdot\frac{\partial y_k}{\partial z_k}=\frac{\partial E}{\partial y_k}\cdot g'(z_k)$$

(b) $$\frac{\partial E}{\partial z_j}=\frac{\partial E}{\partial y_j}\cdot\frac{\partial y_j}{\partial z_j}=\frac{\partial E}{\partial y_j}\cdot g'(z_j)$$

also, $$\frac{\partial E}{\partial y_j}=\sum_k\frac{\partial E}{\partial z_k}\cdot\frac{\partial z_k}{\partial y_j}=\sum_k w_{jk}\frac{\partial E}{\partial z_k}=\sum_k w_{jk}\frac{\partial E}{\partial y_k}g'(z_k)$$

$$\Rightarrow \frac{\partial E}{\partial z_j}=g'(z_g)\cdot\sum_k w_{jk}\cdot g'(z_k)\frac{\partial E}{\partial y_k}$$.#

(c) $$\frac{\partial E}{\partial w_{ij}}=\frac{\partial E}{\partial z_j}\cdot\frac{\partial z_j}{\partial w_{ij}}=\frac{\partial E}{\partial z_j}y_i=g(z_i)\cdot g'(z_j)\sum_k w_{jk}g'(z_k)\frac{\partial E}{\partial y_k}$$ #