Homework 2 Report Problem Set

Professor Pei-Yuan Wu
EE5184 - Machine Learning

電子所碩一 R07943004 莊育權

Problem 1. (1%) 請簡單描述你實作之 logistic regression 以及 generative model 於此 task 的表現，並試著討論可能原因。

	Generative Model	Logistic Regression
準確率(public score)	0.8200	0.8200

上表所用的特徵都是只用 PAY_0，有做標準化，但沒有做 one-hot encoding 出來的結果。可以看出出來的準確率是都一樣的，可能是因為 Generative Model 所假設資料有高斯分布的特性，剛好符合真實資料，使得 Logistic Regression 收斂出來的參數剛好跟 Generative Model 的出來差不多，因此才有此結果。

Problem 2. (1%) 請試著將 input feature 中的 gender, education, marital status 等改為 one-hot encoding 進行 training process，比較其模型準確率及其可能影響原因。

這部分我將沒有做 one-hot encoding 的 gender, education, marriage, pay_0, pay_2, pay_3, pay_4, pay_5, pay_6，九個特徵丟進去訓練，並與這九個特徵去做 one-hot encoding 出來的準確率去做比較。

	無 one-hot	有 one-hot
準確率(public score)	0.8080	0.8215

從上表可以得知有做 one-hot encoding 出來的準確率較高，可能是因為機器學習中的模型大部分都是用空間當中的距離去做分類，因此如果分類資料用 1,2,3 來表示第一類、第二類和第三類，照理來講，這三類的對於原點距離應該都一樣(都是 1 或 0)，但是第三類的數值因為比第一類的數值還來的大，對於分類就會造成不平均的影響，因此通常有類別的資料用 One-hot encoding 去做分類的效果會較好。

Problem 3. (1%) 請試著討論哪些 input features 的影響較大（實驗方法沒有特別限制，但請簡單闡述實驗方法）。

將全部的特徵進去訓練，並最後觀察每個特徵所佔的權重比，此特徵皆已經過標準化，可以得到下表。

Term	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	**PAY_0**
|Weight|	-0.32147048	-0.09281303	-0.09041041	-0.13131521	0.0583905	**1.09695679**
Term	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1

|Weight|	0.29444746	0.10733546	0.00360906	0.04404544	0.08401164	-0.04104944
Term	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_AMT1
|Weight|	-0.03456894	0.02512787	0.06330566	-0.02495179	-0.02495179	-0.12106294
Term	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6	
|Weight|	-0.17289853	0.00395119	-0.0391847	-0.0820094	-0.06632047	

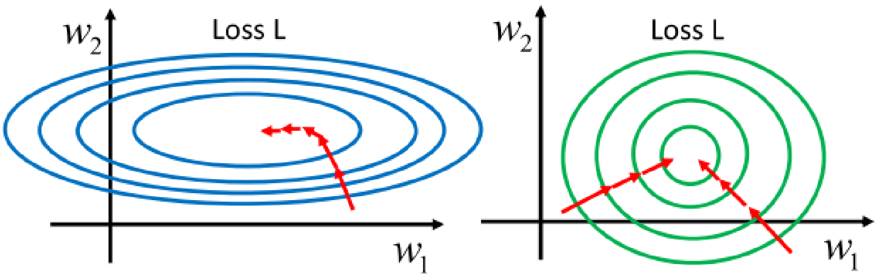
由上表可以得知，PAY_0 所佔的權重最大，因此我將 PAY_0 單獨去做訓練模型，並去比較原本全部特徵的準確率。由下表可以看出單獨只運用 PAY_0 這個特徵，可以提升整題的準確率，可能是因為 PAY_0 的影響較大，而其他的特徵對於預測來說是不太必要，較為冗贅，加進去反而會拖累整體準確率。

	全部特徵有標準化	PAY_0 有做標準化
準確率(public score)	0. 8073	0.8200

Problem 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。

	全部特徵沒有標準化	全部特徵有做 z-score
準確率(public score)	0.7915	0.8073

由實驗結果可以得知，有標準化的特徵做出來的準確率會較高。這個結果可能是因為沒有做標準化的特徵在做訓練的時候，會受特別大的數值所影響，導致整個 loss 的移動不易往最小路線走，也就是不容易收斂，所以準確率較低，如上課圖所示。



Problem 5. (1%) The Normal (or Gaussian) Distribution is a very common continuous probability distribution. Given the PDF of such distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

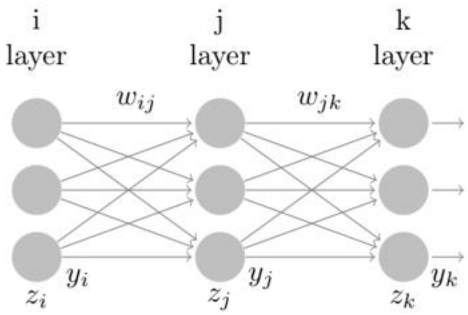
Please show that such integral over $(-\infty, \infty)$ is equal to 1.

Ans:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = A \quad (z = \frac{x-\mu}{\sigma})$$

$$A^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dx\, dy$$
$$(Polar\ transform: x = rcos\theta, y = rsin\theta, dxdy = rdrd\theta)$$
$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} rdr\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left[-e^{-\frac{r^2}{2}}\right]_0^{\infty} d\theta = \frac{1}{2\pi} \int_0^{2\pi} \big(0 - (-1)\big)\, d\theta = 1$$
$$we\ found\ that\ A^2 = 1, and\ we\ know\ A > 0, therefore\ A = 1\ \ => \ Q.E.D.$$

Problem 6. (1%) Given a three layers neural network, each layer labeled by its respective index variable. I.e. the letter of the index indicates which layer the symbol corresponds to.



For convenience, we may consider only one training example and ignore the bias term. Forward propagation of the input zi is done as follows. Where g(z) is some differentiable function (e.g. the logistic function).

$$y_i = g(z_i)$$
$$z_j = \sum_i w_{ij} y_i$$
$$y_j = g(z_j)$$
$$z_k = \sum_j w_{jk} y_j$$
$$y_k = g(z_k)$$

Derive the general expressions for the following partial derivatives of an error function E, also some differentiable function, in the feed-forward neural network depicted. In other words, you should derive these partial derivatives into "computable derivative"

$$(a)\frac{\partial E}{\partial z_k} \quad (b)\frac{\partial E}{\partial z_j} \quad (c)\frac{\partial E}{\partial w_{ij}}$$

(a)

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

(b)

$$\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial z_k} = g'(z_j)\left(\frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial y_j}\right) = g'(z_j)(w_{jk})\left(\frac{\partial E}{\partial y_k}\,g'(z_k)\right)$$

(c)

$$\frac{\partial E}{\partial w_{i,j}} = \frac{\partial E}{\partial z_j}\frac{\partial z_j}{\partial w_{i,j}} = \left(g'(z_j)g'(z_k)\,\frac{\partial E}{\partial y_k}\,w_{jk}\right)(y_j)$$