Homework 2 Report - Credit Card Default Payment Prediction

B05901022 電機三 許睿洋

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。

從整體表現看來，logistic regression 表現的比 generative model 略好(約高 1% 的 accuracy)，數值大約都落在 81~82%附近。其中，generative model 相較於 logistic regression 少了一些進步的空間，而後者搭配 One hot 後便可以超出前者的準確率。在這次的作業中，資料的資料量不少，而且資料本身可能錯誤率也不大，所以 generative model 的表現會不如 logistic regression。

2. (1%) 請試著將 input feature 中的 gender, education, martial 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。

未做 one-hot encoding: accuracy = 78.45%

有做 one-hot encoding: accuracy = 82.05%

在沒有 one-hot encoding 的幫助下，logistic regression 的結果相當差。這可能是因為某些參數的數值大小並不能很好的表現其變化，譬如教育程度，我們無法從 1~6 的數值中看出他的教育程度有何差別。雖然 one-hot 可能會占用到很多 input 的空間，但是很顯然這個結果是相當值得的。

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)。

Canceled feature	0	1	2	3	4	5	6	7	8	9
accuracy	77.75	82.03	81.94	81.96	77.73	80.3	81.94	81.86	82.11	82.01
Canceled feature	10	11	12	13	14	15	16	17	18	19
accuracy	82.05	77.78	77.78	77.78	77.78	77.78	77.78	77.78	77.78	77.78
Canceled feature	20	21	22	No cancel						
Accuracy	77.78	77.78	77.78	81.76						

實驗方法:將 input data 的某一 column(feature)刪掉，其餘參數、方法保持相同，檢測其結果之準確度。

結論:所有 77.78 應該都是把所有的結果都輸出為 0，而這是不能接受的結果，因此顯然將這些數據拔掉會劣化整個訓練結果。而項目中的 1,2,3,5,6,7,8,9,10 拔掉都有不錯的結果，顯示我們在判斷這些信用卡繳款問題時，性別、教育程度、結婚、過去付款情況沒有太大關係。

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。

未做 feature normalization: accuracy = 76.63%

有做 feature normalization: accuracy = 81.76%

從上一題的結論看來，第一項及最後幾項與結論有相當程度的相關性，而這幾項與持卡人的金額額度有關，這些資料往往都有極大的差距，因此特徵標準化對於結果而言有相當重要的關係，可以將這些極大的數據進行處理方便學習的過程。

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

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

please show that such integral over (−∞,∞) is equal to 1.

Collab: Self

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

let $u = \frac{x-\mu}{\sigma}$, $du = \frac{1}{\sigma} dx$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{-\frac{1}{2}u^2} \cancel{\sigma}\,du = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = I$$

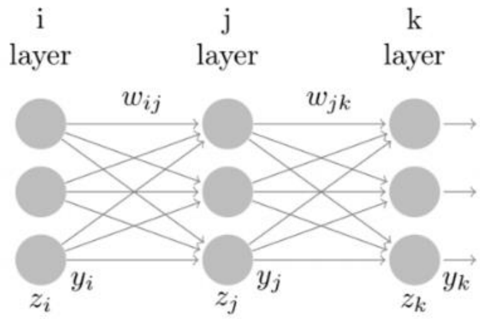
$$I^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}v^2} dv$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(u^2+v^2)} du\,dv$$

$$= \int_{0}^{2\pi} \int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}r^2} r\,dr\,d\theta$$

$$= -(0-1) = 1 \ \#$$

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 z_i is done as follows. Where g(z) is some differentiable function(e.g. the logistic function).

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

Derive the general expressions for the following partial derivatives of an error function E, also sime differentiable function, in the feed-forward neural network depicted. In other words, you should derive these partial derivatives into "computable derivative" (e.g. $\frac{\partial E}{\partial y_k}$ or $\frac{\partial z_k}{\partial w_{jk}}$)

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

Collab: Self

6. (a) $\frac{\partial E}{\partial z_k} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k} = \frac{\partial E}{\partial y_k} g'(z)$

(b) $\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial z_j}$

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

(c) $\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial z_j}\frac{\partial z_j}{\partial w_{ij}}$

$= \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial z_j}\sum_i y_i$