Homework 2 - Credit Card Default Payment Prediction

資工系博士班二年級 D06922023 顏志軒

2018 年 11 月 3 日

這份報告探討 logistic regression 與 generative model 的效能。這份報告中，logistic regression 均使用 cross entropy 作為 loss function，Adagrad 進行參數最佳化，learning rate 為 0.5，迭代次數 epoch 為 10000。

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

我用 min-max feature scaling 將資料作前處理後，分別用 Gaussian generative model 與 logistic regression 訓練，所得模型準確度如下表:

訓練方式	所得模型準確度
min-max feature scaling+generative model	0.49835
min-max feature scaling+logistic regression	0.81115

由表中可知，logistic regression 明顯優於 Gaussian generative model。這是由於 Gaussian generative model 假設每個 feature 都符合高斯分布，而實際上並非如此。

Problem 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。

原始資料有 23 個 feature。進行 one hot encoding 時，我是根據資料中實際有的數值。SEX 並沒有人是 3(其他)，只有 1 與 2 兩種值，EDUCATION 有 06 七種數值，MARRIAGE 有 13 三種數值，原始資料進行 one hot encoding 後具有 32 個 feature。我藉由比較 logistic regression 分別加上 min-max feature scaling+one hot encoding 與只有 min-max feature scaling 兩種訓練方式所得模型的準確度，來量化 one hot encoding 對準確度的影響。結果如下表所示:

訓練方式	所得模型準確度
min-max feature scaling+one hot encoding	0.8114
僅有 min-max feature scaling	0.81115

可以看出, one-hot encoding 後準確度略微上升。在分類型的資料中（例如：性別），不同類型（例如：男、女）對於 Y 值的影響差異並非固定比例，但原始資料中用不同數字來表示不同類別，卻將不同類別的影響固定下來，因此準確度會受到影響。One-hot encoding 避免了這種情況發生，因此能提高準確度。

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

我將未經 one-hot encoding 處理的資料刪除不同列進行 logistic regression 訓練，測量其所得模型的準確度來判斷 feature 影響的程度，準確度下降的量越多，表示這個 feature 影響越大。結果後方如表 1所示，婚姻狀況、教育程度和性別的影響較大。

Problem 4. (1%) 請實作特徵標準化 (feature normalization), 討論其對於你的模型準確率的影響。

我分別使用 logistic regression 對 one hot encoding 後，再對所有 feature 都進行 min-max feature scaling 的資料進行訓練，與單獨使用 one-hot encoding 的資料進行 logistic regression，所得的模型準確度進行比較。結果如下表所示：

訓練方式	所得模型準確度
min-max feature scaling+one hot encoding	0.8114
僅有 one-hot encoding	0.7848

可以看出，min-max feature scaling 大幅增加準確度。這是由於 feature scaling 會使特徵的等高線較為接近 N 維球體 (N 是特徵的數量)，因此收斂速度較快。

另外，單獨使用 one-hot encoding 處理的資料進行 logistic regression，numpy 丟出 `overflow encountered in exp`、`divide by zero encountered in log`、`invalid value encountered in multiply` 等警告訊息。這是由於某些欄位 (例如：信用額度) 數值太大，計算的過程中有超出浮點數範圍的數值。feature scaling 將 feature 數值縮放到 0I 之間，避免的這種情況發生，也是準確度增加的原因。

Problem 5. (1%) The Normal (or Gaussian) Distribution is a very

表 1: 移除不同 feature 對於模型準確度的影響

移除的 feature	所得模型準確度
LIMIT_BAL	0.81055
SEX	0.81025
EDUCATION	0.8104
MARRIAGE	0.81015
AGE	0.8106
PAY_0	0.81085
PAY_2	0.8109
PAY_3	0.8109
PAY_4	0.811
PAY_5	0.81075
PAY_6	0.81055
BILL_AMT1	0.81045
BILL_AMT2	0.81045
BILL_AMT3	0.81045
BILL_AMT4	0.81045
BILL_AMT5	0.81045
BILL_AMT6	0.81045
PAY_AMT1	0.81045
PAY_AMT2	0.81045
PAY_AMT3	0.81045
PAY_AMT4	0.81045
PAY_AMT5	0.81045
PAY_AMT6	0.81045

common continuous probability distribution. Given the PDF of such distribution

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

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

所求的值為:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{2}$$

令此值為 I。此時 I^2 的值為:

$$\begin{aligned} I^2 = &\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right)^2 \\ &\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \right) \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \right) \\ = &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} e^{-\frac{(x-\mu)^2+(y-\mu)^2}{2\sigma^2}} dxdy \end{aligned} \tag{3}$$

令 $x = \mu + \sigma r \cos\theta, y = \mu + \sigma r \sin\theta$,則 $dxdy = \sigma^2 r dr d\theta$,方程式 3可變換為:

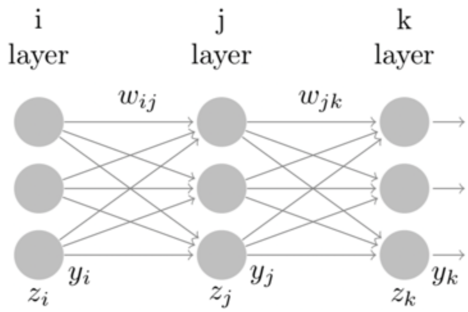
$$I^2 = \int_0^{2\pi} \int_0^{\infty} \frac{1}{2\pi} e^{-\frac{r^2}{2}} r dr d\theta \tag{4}$$

接著,令 $t = e^{-\frac{r^2}{2}}$,則 $dt = -re^{-\frac{r^2}{2}} dr$,方程式 4成為:

$$\begin{aligned} I^2 &= \int_0^{2\pi} -\frac{1}{2\pi} e^{-\frac{r^2}{2}} \Big|_0^{\infty} d\theta \\ &= \int_0^{2\pi} \frac{1}{2\pi} d\theta \\ &= \frac{\theta}{2\pi} \Big|_0^{2\pi} \\ &= 1 \end{aligned} \tag{5}$$

由於 $f(x)$ 在整個實數軸皆為正數,因此 I 也是正數,即為 1。

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

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

Derive the general expressions for the following partial derivatives of an error function E, also 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) $\frac{\partial E}{\partial w_{ij}}$

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