Homework 2 Report

Professor Pei-Yuan Wu

EE5184 - Machine Learning

學號：R07922135 系級：資工研一 姓名：顏百謙

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。

Logistic regression 以及 generative model 都是選擇處理過的 PAY_0 ~ PAY_6 以及對其做 one-hot encoding 的feature進行training.

其中 Logistic regression 的參數如下：

- learning rate = 0.0001
- batch = 50
- iteration = 10000

得到的結果 generative model 較 logistic regression 優異 （左private 右public）（上logistic 下generative)

test1026_logistic_ver1.csv 3 days ago by r07922135_omuraisu add submission details	0.81380	0.81300	☐
test1026_ver6.csv 6 days ago by r07922135_omuraisu ALL PAY	0.82300	0.82160	✔

其原因有可能為：

- 資料量不足
- iteration次數不夠高
- 運氣不好

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

- 以generative model進行預測可得到

- One-hot encoding & Without one-hot encoding

homework2_no_one_hot.csv just now by r07922135_omuraisu add submission details	0.80620	0.81200	☐
homework2_one_hot.csv 2 minutes ago by r07922135_omuraisu one_hot	0.80800	0.80920	☐

由Private的結果來看，進行one_hot後會有較好的結果。

其原因可能為教育層度/結婚中的feature值並該不含有線性相關的特徵（例如：教育層度每個level並不該是等差）

若沒有做one hotu就下去train可能會不小心train進了不該有的關係。

因此當在train此類數值間沒有等距相對關係的feature時用one hot可以達到較好的效果。

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

homework2_all_without_pay.csv just now by r07922135_omuraisu add submission details	0.78120	0.78060	☐
homework2_all_without_sex_edu_marry_age.csv 3 minutes ago by r07922135_omuraisu without	0.80800	0.81140	☐
homework2_all.csv 4 minutes ago by r07922135_omuraisu all2	0.80620	0.81200	☐
homework2_pay_only_with_one_hot.csv 7 minutes ago by r07922135_omuraisu pay_only_with_one_hot	0.82300	0.82160	☐
homework2_pay_only.csv 7 minutes ago by r07922135_omuraisu pay_only	0.80600	0.80920	☐

以generative model進行了5種feature組合的training得到了以上結果，

其中可看出幾種特徵：

- 由第二個和第三個submission可看出 性別 教育程度 結婚 年齡 等feature對結果的影響並不顯著
- 由第一個和第五個submission可看出 pay_0 ~ pay_6的feature影響層度佔了非常大的比重
  - 僅以 pay_0 ~ pay_6 進行training得到的結果和用所有feature train的差不多
  - 拿掉pay_0 ~ pay_6後準確度大幅下降
- 由第四個submission可看出 pay_0 ~ pay_6 佔了很大的比重，而且還可對其進行feature的優化，以達到更好的結果。

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。

- generative model (feature: 對 pay_0 ~ pay_6 進行 one-hot encoding)
  - normalization

homework2_4_normalization.csv 6 hours ago by r07922135_omuraisu normalization	0.82240	0.82120	☐

  - Without normalization

homework2_4_without_normalization.csv 6 hours ago by r07922135_omuraisu without normalization	0.82240	0.82120	☐

結果顯示，在generative model下是否normalization對結果沒有影響。

原因可能是因為normalize只是對feature進行等比例縮放，並不會對generative model的生成方式產生影響

- Logistic model (feature: 所有feature進行 training)
  - normalization

homework2_4_logistic_with_normalize.csv a few seconds ago by r07922135_omuraisu add submission details	0.79600	0.79740	☐

  - Without normalization

```
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site-packages/ipykernel_launcher.py:4: RuntimeWarnin
g: overflow encountered in exp
  after removing the cwd from sys.path.
[81879.92606239] 0
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site-packages/ipykernel_launcher.py:7: RuntimeWarnin
g: overflow encountered in exp
  import sys
/Library/Frameworks/Python.framework/Versions/3.7/lib/python3.7/site-packages/ipykernel_launcher.py:13: RuntimeWarnin
g: overflow encountered in exp
  del sys.path[0]
[157767.04028085] 100
```

結果顯示在沒有normalization的狀況下training過程很容易overflow進而造成model產生問題。

而且在沒有進行normalization的時候gradient descent的方向可能會特別偏向某些feature而淡化掉其他feature的影響力。

Problem 5.

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.

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

設 $(x-\mu) = u$ 且 $du = dx$

則上式 $= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(u)^2}{2\sigma^2}} du$

$$let \int_{-\infty}^{\infty} e^{\frac{-(u)^2}{2\sigma^2}} du = \int_{-\infty}^{\infty} e^{\frac{-(v)^2}{2\sigma^2}} dv = I$$

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{\frac{-r^2}{2\sigma^2}} rdrd\theta$$

$$= \int_0^{2\pi} d\theta \int_0^{\infty} e^{\frac{-r^2}{2\sigma^2}} rdr$$

$$= 2\pi(-\sigma^2 e^{\frac{-r^2}{2\sigma^2}} |_0^{\infty})$$

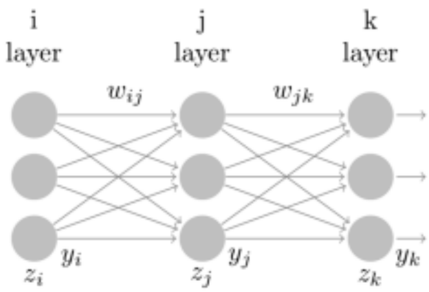
$$= 2\pi(0 - (-\sigma^2))$$

$$= 2\pi\sigma^2$$

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

Problem 6.

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).

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

Derive the general expressions for the following partial derivatives of an error function E in the feed-forward neural network depicted.

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

(a)設 $\frac{\partial E}{\partial z_k}$ 為求 E對第$k\ layer$整層的 $z_{k=n}$ 做偏微分的向量值

我們計算其中一層（第a層）的結果

$$
\begin{aligned}
\frac{\partial E}{\partial z_{k=a}} &= \frac{\partial y_a}{\partial z_{k=a}} \frac{\partial E}{\partial y_a} \\
&= g'(z_{k=a}) \frac{\partial E}{\partial y_a}
\end{aligned}
$$

故可推出 $\frac{\partial E}{\partial z_k} = \begin{bmatrix} g'(z_{k=1}) \frac{\partial E}{\partial y_1} \\ g'(z_{k=2}) \frac{\partial E}{\partial y_2} \\ \vdots \\ g'(z_{k=n}) \frac{\partial E}{\partial y_n} \end{bmatrix}$

(b)設 $\dfrac{\partial E}{\partial z_j}$ 為求 E 對第 j $layer$ 整層的 $z_{j=m}$ 做偏微分的向量值

我們計算其中一層（第 a 層）的結果

$\Big($設第 k $layer$ 有 n 層，且第 j $layer$ 的第 q 層到第 k $layer$ 第 r 層的 w 寫為 $w_{ij=qr}\Big)$

$$\frac{\partial E}{\partial z_{j=a}} = \sigma'(z_{j=a})[\sum_{r=1}^{n} w_{ij=ar}\frac{\partial E}{\partial z_{k=r}}]$$

故可推出 $$\frac{\partial E}{\partial z_j} = \begin{bmatrix} \sigma'(z_{j=1})[\sum_{r=1}^{n} w_{ij=1r}\frac{\partial E}{\partial z_{k=r}}] \\ \sigma'(z_{j=2})[\sum_{r=1}^{n} w_{ij=2r}\frac{\partial E}{\partial z_{k=r}}] \\ \vdots \\ \sigma'(z_{j=m})[\sum_{r=1}^{n} w_{ij=mr}\frac{\partial E}{\partial z_{k=r}}] \end{bmatrix}$$

(c)設 $\dfrac{\partial E}{\partial w_{ij}}$ 為求 E 對第 i $layer$ 及第 j $layer$ 間的其中一個 $w_{ij=qs}$ 做偏微分的結果

則 $$\frac{\partial E}{\partial w_{ij=qs}} = \frac{\partial z_{j=r}}{\partial w_{ij=qs}}\frac{\partial E}{\partial z_{j=s}}$$

$$= (z_{i=q})\frac{\partial E}{\partial z_{j=s}}$$

$$= (z_{i=q})\sigma'(z_{j=s})[\sum_{r=1}^{n} w_{ij=sr}\frac{\partial E}{\partial z_{k=r}}]$$

若要擴展至所有 i $layer$ 和 j $layer$ 間的 w_{ij} $\big($設第 i $layer$ 共 l 層$\big)$則

$$\frac{\partial E}{\partial w_{ij}} = \begin{bmatrix} (z_{i=1})\sigma'(z_{j=1})[\sum_{r=1}^{n} w_{ij=1r}\frac{\partial E}{\partial z_{k=r}}] \\ (z_{i=1})\sigma'(z_{j=2})[\sum_{r=1}^{n} w_{ij=2r}\frac{\partial E}{\partial z_{k=r}}] \\ \vdots \\ (z_{i=1})\sigma'(z_{j=m})[\sum_{r=1}^{n} w_{ij=mr}\frac{\partial E}{\partial z_{k=r}}] \\ (z_{i=2})\sigma'(z_{j=1})[\sum_{r=1}^{n} w_{ij=1r}\frac{\partial E}{\partial z_{k=r}}] \\ (z_{i=2})\sigma'(z_{j=2})[\sum_{r=1}^{n} w_{ij=2r}\frac{\partial E}{\partial z_{k=r}}] \\ \vdots \\ (z_{i=l})\sigma'(z_{j=m})[\sum_{r=1}^{n} w_{ij=mr}\frac{\partial E}{\partial z_{k=r}}] \end{bmatrix}$$