Homework 2 Report

Professor Pei-Yuan Wu

EE5184 - Machine Learning

學生：洪正皇　學號：r07922050

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。

在沒有實作 one-hot encoding 及 feature normalization 之前，logistic regression 與 generative model 的分數分別為 0.73320 及 0.78120，在 logistic regression 的部分我認為是不重要的 features 嚴重影響了結果，而使 model 不僅分類不出正確的 class1，還將很多 class0 分成 class1，導致結果很差；而 generative model 也是受到數字分佈較大的 model 影響，Gaussian distribution 中的 sigma 的行列式值過大，使得機率都接近 0。我想，generative model 要各個 features 都有滿高機率為 class1 的時候，才會成功分類為 class1，因此沒有大量將 class0 分成 class1 的情況。

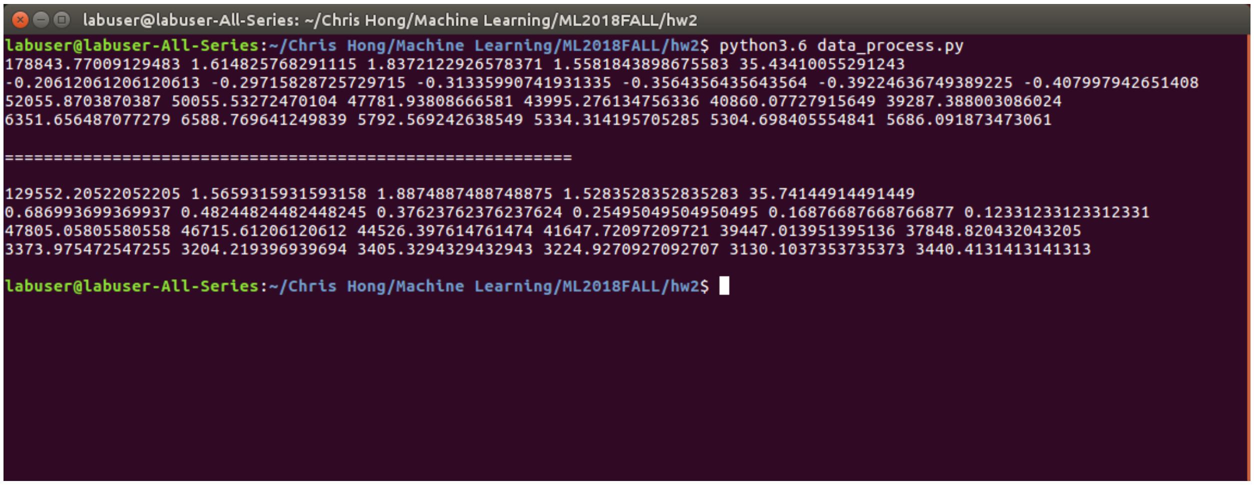
Problem 2. (1%) 請試著將 input feature 中的 gender, education, martial status 等改 one-hot encoding 進行 training process，比較其模型準確率及其可能影響原因。

這些 features 數字大小沒有特別關聯，只是各自代表不同的情況而已，one-hot encoding 能將各個情況分開來，計算他們對 output 的權重與影響。在這次預測中，gender, education, martial status 在 class1 及 class0 中的比例都差不多(依照平均值)，因此實作 one-hot encoding 前後結果都差不多。

model	pure	one-hot encoding	normalization	one-hot encoding + normalization
public score	0.73860	0.73860	0.81580	0.81440
private score	0.73320	0.73320	0.81140	0.81220

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

我將各 class1 與 class0 的資料分開來並分別計算其各項 features 的平均值，其中差異最大的是 x[5:10]、x[17:22]這十二個 features，而後我只針對這些 features 去 train 我的 model，得到了滿大幅度的提升。我也針對多種其他組合的 features 測試，預測結果都與使用全部 features 時差不多。因此我認為 x[5:10]、x[17:22]這十二個 features 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。

在做 feature normalization 之前 x[0]及 x[11:22]的值皆較大，在 logistic regression 中對於 wx+b 項佔有很大的影響比例，使得 sigmoid 的值都很極端；而在 generative model 中，這些值大權重的也影響了 sigma，使其他 features 難以作用。在 normalization 過後，各項 features 都能對結果有影響，而得到比較好的結果，若每個 model 的 features 都 normalization 到 0~1 之間，那 learning rate 也比較好調整。

model	pure	normalization	one-hot encoding	one-hot encoding + normalization
public score	0.73860	0.81580	0. 73860	0.81440
private score	0.73320	0. 81140	0.73320	0.81220

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 (*1; 1*) is equal to 1.

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

$$\boxed{\text{Let } z = \frac{x-\mu}{\sigma}, \quad \frac{dz}{dx} = \frac{1}{\sigma} \Rightarrow dx = \sigma\, dz}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-z^2}{2}} \sigma\, dz$$

$$I^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz\right) \qquad \boxed{\text{Let } p = z}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{-(z^2+p^2)}{2}} dz\, dp \qquad \boxed{\text{Let } r^2 = z^2+p^2, \ 0 \le r \infty \quad z = r\cos\theta,\ p = r\sin\theta,\ 0 \le \theta \le 2\pi}$$

$$= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} e^{\frac{-(r^2\cos^2\theta + r^2\sin^2\theta)}{2}} \left|\begin{vmatrix} \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial r} \\ \frac{\partial p}{\partial \theta} & \frac{\partial p}{\partial r} \end{vmatrix}\right| dr\, d\theta$$

$$\boxed{\left|\begin{vmatrix} \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial r} \\ \frac{\partial p}{\partial \theta} & \frac{\partial p}{\partial r} \end{vmatrix}\right| = \left|\begin{vmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{vmatrix}\right| = |r|\left|-\sin^2\theta - \cos^2\theta\right| = r}$$

$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{\frac{-r^2}{2}} \cdot r\, dr\, d\theta \qquad \boxed{\text{Let } w = r^2, \ \frac{dw}{dr} = 2r \Rightarrow \frac{dw}{2} = r\, dr}$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{\frac{-w}{2}} \cdot \frac{1}{2}\, dw\, d\theta$$

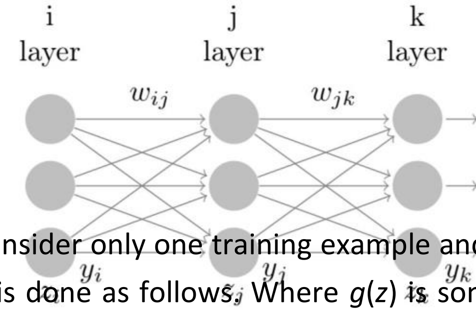
$$= \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{2}\left[-2e^{\frac{-w}{2}}\Big|_0^{\infty}\right] d\theta \qquad \Rightarrow \qquad I = \sqrt{1}$$

$$= \frac{1}{2\pi}\int_0^{2\pi} -[0-1]\, d\theta \qquad = 1 \ \#$$

$$= \frac{1}{2\pi}\int_0^{2\pi} 1\, d\theta$$

$$= 1$$

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).

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

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

(b) $$\frac{\partial E}{\partial z_j} = \frac{\partial g(z_j)}{\partial z_j} \cdot \frac{\partial E}{\partial y_j}$$
$$= g'(z_j) \cdot \sum_k w_{jk} \cdot \frac{\partial E}{\partial z_k}$$
$$= \underline{g'(z_j) \cdot \sum_k \left(w_{jk} \cdot g'(z_k) \cdot \frac{\partial E}{\partial y_k} \right)} \ \#$$

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