Homework 2 Report

資工所碩二 R06922132 何羿辰

EE5184 - Machine Learning

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。

Score	Private	Public
generative model	0.82220	0.82100
logistic regression	0.78120	0.78060

從結果來看 generative model 不論在 Private 或 Public 的分數都比較高，推測可能原因是因為我沒有對 feature 進一步做 feature transform，導致 logistic regression 無法很清楚劃分出邊界。

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

使用 generative model 來做有無 one-hot encoding 的比較

One-hot encoding	Private	Public
Yes	0.82180	0.82020
No	0.80880	0.81040

從結果可以看到使用 one-hot encoding 明顯比較好，原因是因為 gender/ education/ martial sstatus 這些 feature 代表的意義是一個狀態而不應該是數值，因此應該先做轉換再使用。

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

做了幾組嘗試，實驗方法為刪減 feature，最後發現使用 History of past payment 可得到最佳效果，而 Amount of bill statement 與 Amount of previous payment 則會造成較差結果，思考這幾個 feature 的本質，我推測 default payment status 跟客戶的過去信用與還款習慣有關，而 History of past payment 就是整合了 Amount of bill statement 與 Amount of previous payment 所得到的結論，而後兩個 feature 卻會因為金額大小影響訓練的準確度，除非進一步做 feature transform 才能拿來使用。最後附上把 Amount of bill statement 與 Amount of previous payment 刪除掉的結果比較。

Generative model	Private	Public
Delete last 2 features	0.82220	0.82100
All feature	0.82180	0.82020

Problem 4. (1%) 請實作特徵標准化 (feature normalization)，討論其對於你的模型准確率的影響。

比較有無做特徵標準化的結果

Normalization	Private	Public
Yes	0.80660	0.80740
No	0.78120	0.78080

從結果可以看到對 LIMIT_BAL/age/Amount of bill statement/Amount of previous payment 這些非狀態 feature 做 normalization 之後分數會提高很多，我認為這是因為 normalization 可以加快收斂的速度，而避免一些極端值對模型造成過大影響。

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.

取積分:

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

令 $\omega = \frac{x-\mu}{\sigma\sqrt{2}}$ ➔ $d\omega = \frac{dx}{\sigma\sqrt{2}}$ 代入原式➔ $\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} e^{-\omega^2} d\omega$

By Gaussian integral:

$I(a) = \int_{-a}^{a} e^{-x^2} dx$ ➔ $\lim_{a\to\infty} I(a) = \int_{-\infty}^{\infty} e^{-x^2} dx$

➔ $I^2(a) = \left(\int_{-a}^{a} e^{-x^2} dx\right)\left(\int_{-a}^{a} e^{-y^2} dy\right) = \iint e^{-(x^2+y^2)} dxdy$

➔ $\int_0^{2\pi}\int_0^{a} re^{-r^2} drd\theta < I^2(a) < \int_0^{2\pi}\int_0^{a\sqrt{2}} re^{-r^2} drd\theta$

➔ $\pi(1-e^{-a^2}) < I^2(a) < \pi(1-e^{-2a^2})$

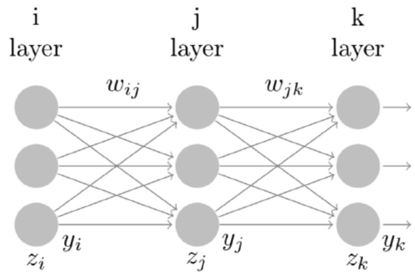
$\lim_{a\to\infty} \pi(1-e^{-a^2}) = \lim_{a\to\infty} \pi\left(1-e^{-2a^2}\right) = \pi$

根據夾擠定理: $I^2(\infty) = \pi$

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

代入原式➔ $\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{\pi}} * \sqrt{\pi} = 1$ 得證#

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) $\frac{\partial E}{\partial z_k} = \frac{\partial}{\partial z_k}\frac{1}{2}\|\overline{y_k} - y_k\|^2 = \frac{\partial}{\partial z_k}\frac{1}{2}\Sigma(\overline{y_k} - g_k(z_k))^2 = -(\overline{y_k} - g_k(z_k)) * g_K{}'(z_k)$

(b) $\frac{\partial E}{\partial z_j} = \frac{\partial}{\partial z_j}\frac{1}{2}\|\overline{y_k} - y_k\|^2 = \frac{\partial}{\partial z_j}\frac{1}{2}\Sigma\left(\overline{y_k} - g_k(z_k)\right)^2$

$= \Sigma -\left(\overline{y_k} - g_k(z_k)\right) * \frac{\partial}{\partial z_j} g_j(z_k) = \Sigma -\left(\overline{y_k} - g(z_k)\right) * g_k{}'(z_k)\frac{\partial z_k}{\partial z_j}$

$= \Sigma \frac{\partial E}{\partial z_k}\frac{\partial}{\partial z_j}\Sigma_j(\omega_{jk} y_j) = \Sigma \frac{\partial E}{\partial z_k}\frac{\partial}{\partial z_j}\Sigma_j(\omega_{jk} * g_j(z_j)) = (\Sigma \frac{\partial E}{\partial z_k} * \omega_{jk}) g_j{}'(z_j)$

(c) $\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial z_j}\frac{\partial z_j}{\partial \omega_{ij}} = \left(\Sigma \frac{\partial E}{\partial z_k} * \omega_{jk}\right) g_j{}'(z_j) * \frac{\partial}{\partial \omega_{ij}}\left(\Sigma_i \omega_{ij} y_i\right)$

$= \left(\Sigma \frac{\partial E}{\partial z_k} * \omega_{jk}\right) g_j{}'(z_j) * y_i$