Homework 2 Report -
Credit Card Default Payment Prediction

資工四 B04902131 黃郁凱

November 2, 2018

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。

   從表格1可以看到，在 validation 結果上，Logistic Regression 有比較少的 overfit 的現象；在 testing 上可以觀察到 Generative Model 有稍微好一些，但基本上兩者的 performance 差不多。Generative Model 目的想找出資料分布，根據機率模型的假設，找出最符合 $P(X, Y)$ 的 μ 跟 Σ；而 Discriminative Model 則是想要直接找出 $P(Y|X)$。上課時老師有提到，大多的情況下 Discriminative Model 的成效會比較好，也就是 Logistic Regression 的表現照理會比較好，但是在 training data 較少的情況下，Generative Model 的 performance 會相對好一些，因為他多了一些機率假設，因此在資料少的情況下，如果資料分布符合那樣的假設，就會發揮較好的作用。在這次的競賽題中，兩萬筆的資料尚無法看出兩個 model 的明顯好壞，如果今天有二十萬筆 training data，那麼我相信 Logistic Regression 會明顯勝過 Generative Model。

Logistic Regression				Generative Model			
train	validation	public	private	train	validation	public	private
0.821	0.821	0.819	0.8216	0.823	0.815	0.821	0.8224

Table 1: 不同模型的正確率比較。

2. 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。

   表格2中可以明顯看出，有做 one-hot encoding 的效果會比較好，不論在 training 還是 testing 的結果都是。在資料當中，那些轉換為 one-hot 的維度，本身數字的值並沒有代表性，只能區隔不同類別的屬性，例如男生是 1 女生是 0。因此，將這些維度轉換為 one-hot encoding 是合理的，也讓模型比較好認出這些屬性的差異。

3. 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)。

   從 Logistic Regression 的實驗結果可以發現，當我先用全部的特徵下去訓練，將訓練好的參數倒出來看，可以發現其中 [PAY_0-6] 這些特徵所佔有的比重較大，

model	Training		Testing	
	train	validation	public	private
有 one-hot	0.820	0.819	0.819	0.822
無 one-hot	0.811	0.814	0.807	0.807

Table 2: 不同模型的正確率比較。

而其他的特徵大多接近 0。為了確保這些特徵是有用的，只用這幾種特徵下去訓練，所得到的正確率不會有太大變化，都約 0.82，更顯示這些特徵的重要性。

4. 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。

   當我沒有加特徵標準化時，訓練過程會很不穩定，並且正確率會一直卡在 60% 左右無法上升，然而加上以後，可以使訓練變穩定且正確值可以收斂到 82%。當我將同個維度的多次方也加進去一起訓練時，這樣的情況會加劇，原因出在沒有標準化的資料如果有巨大變化的值，就會影響線性模型 $\sum w_i x_i$ 的結果，而加上標準化會使得這樣的偏差影響減小，較不會受到 bias 的影響。

5. 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.

   - First prove that

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

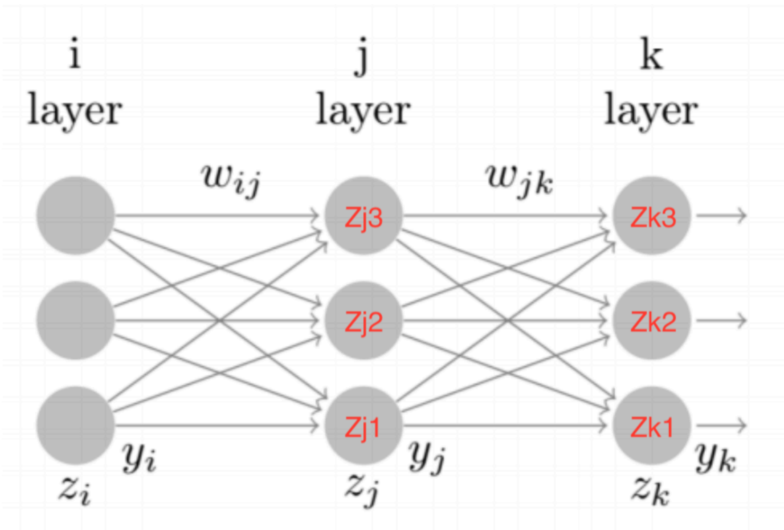
$$\begin{aligned}
I^2 &= 4\int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dy\, dx \\
&= 4\int_0^\infty \left(\int_0^\infty e^{-(x^2+y^2)}\, dy\right)\, dx \\
&= 4\int_0^\infty \left(\int_0^\infty e^{-x^2(1+s^2)} x\, ds\right)\, dx \\
&= 4\int_0^\infty \left(\int_0^\infty e^{-x^2(1+s^2)} x\, dx\right)\, ds \\
&= 4\int_0^\infty \left[\frac{1}{-2(1+s^2)} e^{-x^2(1+s^2)}\right]_{x=0}^{x=\infty}\, ds \\
&= 4\left(\frac{1}{2}\int_0^\infty \frac{ds}{1+s^2}\right) \\
&= 2\Big[\arctan s\Big]_0^\infty \\
&= \pi
\end{aligned}$$

Thus, $I = \sqrt{\pi}$ (Prove from wiki:`https://en.wikipedia.org/wiki/Gaussian_integral`)

- Then prove the required integral is 1.

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(\frac{x-\mu}{\sqrt{2}\sigma})^2}\,dx \\
\text{let } z &= \frac{x-\mu}{\sqrt{2}\sigma} \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-z^2}\sqrt{2}\sigma\,dz \\
&= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2}\,dz \\
&= \frac{1}{\sqrt{\pi}} I \\
&= \frac{1}{\sqrt{\pi}} \sqrt{\pi} \\
&= 1
\end{aligned}
$$

6. 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, also sime differentiable function, in the feed-forward neural network depicted. In other words, you should derive these partial derivatives into "computable derivative"

(a) 首先，我將 z_k 那層的三個神經元分別標上 z_{k1}、z_{k2} 和 z_{k3}，而 y_k 也是。

$$
\begin{aligned}
\frac{\partial E}{\partial z_{ka}} &= \frac{\partial E}{\partial y_{ka}} \frac{\partial y_{ka}}{\partial z_{ka}} \\
&= \frac{\partial E}{\partial y_{ka}} \frac{\partial g(z_{ka})}{\partial z_{ka}} \\
&= \frac{\partial E}{\partial y_{ka}} g'(z_{ka})
\end{aligned}
$$

, for $a \in \{1, 2, 3\}$

(b) 將 z_j 那層的三個神經元分別標上 z_{j1}、z_{j2} 和 z_{j3}，而從 z_{ja} 對應到 z_{kb} 的參數是 $w_{j_a k_b}$，當中的 $a, b \in \{1, 2, 3\}$。

$$
\begin{aligned}
\frac{\partial E}{\partial z_{ja}} &= \left(\sum_{b=1}^{3} \frac{\partial E}{\partial y_{kb}} \frac{\partial y_{kb}}{\partial z_{kb}} \frac{\partial z_{kb}}{\partial y_{ja}} \right) \frac{\partial y_{ja}}{\partial z_{ja}} \\
&= \left(\sum_{b=1}^{3} \frac{\partial E}{\partial y_{kb}} g'(z_{kb}) w_{j_a k_b} \right) g'(z_{ja})
\end{aligned}
$$

, for $a \in \{1, 2, 3\}$

(c) 從 z_{ia} 對應到 z_{jb} 的參數是 $w_{i_a j_b}$，從 z_{jb} 對應到 z_{kc} 的參數是 $w_{j_b k_c}$ 當中的 $a, b, c \in \{1, 2, 3\}$。

$$
\begin{aligned}
\frac{\partial E}{\partial w_{i_a j_b}} &= \left(\sum_{c=1}^{3} \frac{\partial E}{\partial y_{kc}} \frac{\partial y_{kc}}{\partial z_{kc}} \frac{\partial z_{kc}}{\partial y_{jb}} \right) \frac{\partial y_{jb}}{\partial z_{jb}} \frac{\partial z_{jb}}{\partial w_{i_a j_b}} \\
&= \left(\sum_{c=1}^{3} \frac{\partial E}{\partial y_{kc}} g'(z_{kc}) w_{j_b k_c} \right) g'(z_{jb}) y_{ia}
\end{aligned}
$$