Homework 2 Report –

Credit Card Default Payment Prediction

學號：R06521504 系級：土木所交通組碩二 姓名: 陳譽仁

Problem 1

本題採用右列的 feature: PAY_0, BILL_AMT1, PAY_AMT1, PAY_AMT2。Logistic regression 使用 adaptive gradient decent，單一 batch，learning rate 取 10，計算 10000 個 epoch，generative model 則參考講義的計算方式。結果如下表：

	Cross Entropy	Public Score	Private Score
Logistic regression	8486.7438	0.81360	0.81180
Generative model	11382.8433	0.78100	0.78360

可以發現 Logistic regression 的表現都比較好，原因可能是 generative model 對 $P(C_1)$、$P(x|C_1)$等機率分布有事先的假設，但資料分布與假設有較大落差，使得計算結果不如只考慮 $P(y|x)$的 logistic regression 來的好。

Problem 2

本題除了 SEX, EDUCATION, MARRIAGE 外，放入所有的 feature，並設計三個實驗，分別為不用 one hot encoding、原 feature 與 one-hot encoding 並存、僅使用 one-hot encoding 的三種設定，但有考量 one-hot encoding 的設定都保留所有類別所生成的 feature。計算上先由測試資料中取出 10%作為自行測試的資料，使用 logistic regression，adaptive gradient decent，單一 batch，learning rate 取 10，計算 10000 個 epoch。結果如下表：

設定	Cross Entropy	Public Score	Private Score
不用 one-hot	8366.9384	0.81780	0.81900
並存	8350.8305	0.81800	0.81920
僅 one-hot	8350.0.538	0.81900	0.81800

從結果可以發現不用 one-hot encoding 略差於考慮 one-hot encoding，但是是否去除原始分類資料對結果沒有很明顯的影響。以 EDUCATION 來說，不用 one-hot encoding 使模式只能考慮學歷越低，Y 值越低或越高，而無法考

慮各類學歷對於結果的影響，因此使用 one-hot encoding 使模式能計算各類別的影響，使模式表現變好。

Problem 3

我覺得影響較大的input feature代表去除該feature會使模式的表現下降最多，因此本題從本次作業在 Kaggle 上最終的設定開始(所使用的 feature 包含所有原有 x 資料的欄位，再加上對 SEX, EDUCATION, MARRIAGE 進行 one hot encoding，並對於所有資料進行 min max 的 feature scaling，再由測試資料中取出 10%作為自行測試的資料)，測試刪除各 feature 對模式的影響，取由 training data 切割出的測試資料中下降前五多的設定(分別是刪除 PAY_0, BILL_AMT1, SEX_male (對 male 進行 one-hot encoding), PAY_AMT5, MARRIAGE_single (對 single 進行 one-hot encoding))，再上傳 Kaggle 觀察分數，下降最多與次多的結果如下：

刪除的 feature	Cross Entropy	Public Score	Private Score
無	8350.8305	0.81800	0.81920
PAY_0	8656.6819	0.80620	0.80020
BILL_AMT1	8358.7121	0.81440	0.81380

由結果來看，因為刪除 PAY_0 與刪除 BILL_AMT1 會使模式表現降低較多，而認為該兩者是較為重要的 input feature。

Problem 4

在測試的過程中發現，如果沒有對資料進行 feature scaling，會因為資料的尺度相差過大(例如 LIMIT_BAL 為數千至數萬、SEX 為 1、2)，造成 gradient decent 時由於特定 feature 的尺度大，對應的 gradient 就很大，使尺度較小 feature 的 weight 對結果的影響很小，讓模式難以考慮到尺度較小的 feature，也容易因為某些 weight 的變化太大而難以收斂，因此本次作業最終在 Kaggle 上的成果使用 min max 的 feature scaling。

Problem 5

Let

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \hat{x} = \frac{x-\mu}{\sqrt{2}\sigma} \Rightarrow d\hat{x} = \frac{1}{\sqrt{2}\sigma} dx \Rightarrow dx = \sqrt{2}\sigma d\hat{x}$$

Also, let

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

$$\therefore I(x) \cdot I(y) = \int_{-\infty}^{\infty} f(x)dx \int_{-\infty}^{\infty} f(y)dy = \frac{1}{\pi} \int_{-\infty}^{\infty} e^{-\hat{x}^2} d\hat{x} \int_{-\infty}^{\infty} e^{-\hat{y}^2} d\hat{y}$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(\hat{x}^2+\hat{y}^2)} d\hat{x}\, d\hat{y}$$

Let $\begin{cases} \hat{x} = r\cos\theta \\ \hat{y} = r\sin\theta \end{cases} \Rightarrow d\hat{x}d\hat{y} = rdrd\theta$,

$$\Rightarrow I(x) \cdot I(y) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2} rdr\, d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{\infty} e^{-r^2} dr^2 \right) d\theta$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \left(-e^{-r^2} \Big|_0^{\infty} \right) d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta = 1$$

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

Problem 6

(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)

$$\frac{\partial E}{\partial z_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial z_j} = \frac{\partial E}{\partial y_j} g'(z_j) = \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial y_j} g'(z_j) = \frac{\partial E}{\partial z_k} \left[\frac{\partial}{\partial y_j} \left(\sum_j w_{jk} y_j \right) \right] g'(z_j)$$

$$= \frac{\partial E}{\partial y_k} g'(z_k) g'(z_j) \sum_j w_{jk}$$

(c)

$$\frac{\partial \mathrm{E}}{\partial \mathrm{w_{ij}}} = \frac{\partial E}{\partial z_j} \frac{\partial z_j}{\partial w_{ij}} = \frac{\partial E}{\partial z_j} \left[\frac{\partial}{\partial w_{ij}} \left(\sum_i w_{ij} y_i \right) \right] = \frac{\partial E}{\partial z_j} \sum_i y_i$$

$$= \frac{\partial E}{\partial y_k} g'(z_k) g'(z_j) \sum_j w_{jk} \sum_i y_i$$