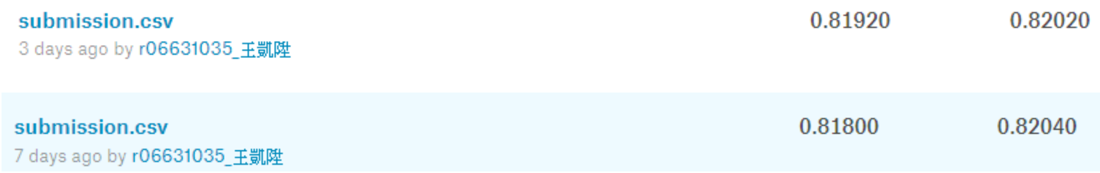
Homework 2 Report

學號：r06631035　系級：生機所碩二　姓名: 王凱陞

1. 以下兩個結果都已做過同樣的資料處理，確保兩者資料一致進行比對。



上圖白底(上方)為 Generative model、藍底(下方)為 Logistic regression 於 Kaggle 所跑出的結果，左為 private score，右為 public score，可看到 Generative model 所得到的 public score 略小於 Logistic regression 所跑出的結果，但 privates score 卻來的好上許多，我認為 Generative model 是利用貝氏定理進行計算，透過了 maximum likelihood 進行高斯分布參數的計算，也許此次的 public 選用的資料較沒有 private 的資料來得符合高斯分布，因此在這部份便會產生些微誤差，而這次的資料集理當比較接近於高斯分布，使得 Generative model 並沒有太差，也或許是在做 Logistic regression 時有些過度的 overfitting 到 public 資料所導致。

2. 以下結果，除了題意要求，其餘資料處理皆保持一致。

圖中上面為根據題意將 input feature 中的 gender, education, martial status 改為 one-hot encoding 所得到的結果，下面則為沒有進行 one-hot encoding 的結果，可看到在 public、private score 有做 one-hot encoding 的分數都較好，因為原先這些 feature 利用數字進行分類，其代表的數字不和要分類的結果一致，如 education 出現 0-6 的數字分類，但是結果只有 0/1，在 training 的時候，計算上這樣的 feature 便需要進行 one-hot encoding 讓 education 分成七種狀況，運用 0/1 進行分辨是否有發生，這樣在 training 的結果才會較為符合。

3.　在第二題中使用 gender, education, martial status 進行 one-hot encoding，會比單獨只使用 education 進行 one-hot encoding 的效果較差一些，下表中提供了一些測試後的數據，接著我將每個單獨的 feature 進行 training，發現單獨使用 PAY_0 就有很好的結果，所以在之後所以在 final submission 的 model 中，選擇將 PAY_0 這項進行二次方和四次方的計算，並當作額外的 feature 加入資料集中，也對結果有了大幅度地提升。

	private	public
PAY_0	0.81900	0.82000
LIMIT_BAL、Martial、education、PAY_0	0.81940	0.82020
LIMIT_BAL、gender、education、PAY_0	0.81520	0.81780
LIMIT_BAL、education、PAY_0	0.81940	0.82040
education、PAY_0	0.81940	0.82040
education、PAY_0-6 (no 1)	0.81980	0.82080
education、PAY_0、BILL_AMT 1-6、PAY_AMT 1-6	0.81680	0.81660

<結論>　在此次的資料中，PAY_0 的影響最大、其餘 PAY_2-6 和做了 one-hot encoding 後的 education 影響為其次，其餘 feature 對結果的影響都甚小，甚至 gender、BILL_AMT 1-6 和 PAY_AMT 1-6 都對結果有不佳的影響。

4. 經過上題的測試，由於 feature 中，LIMIT_Bal、EDUCATION、PAY_0 影響的效果較大，所以僅保留這幾項 feature，但這幾個 feature 僅 LIMIT_BAL 較需要使用 feature normalization，EDUCATION 因為做過 one-hot encoding 影響不大，PAY_0 也因為資料的分布範圍較小影響不大，故在此提多加入 BILL_AMT 1-6 和 PAY_AMT 1-6，並對 LIMIT、BILL_AMT 1-6 和 PAY_AMT 1-6 做特徵標準化探討:

先比對 validation 的狀況:

```
E_val: 0.1886 , epoch: 220
E_val: 0.1886 , epoch: 221
E_val: 0.1884 , epoch: 222
E_val: 0.1882 , epoch: 223
E_val: 0.1882 , epoch: 224
E_val: 0.1882 , epoch: 225
```

```
E_val: 0.7824 , epoch: 220
E_val: 0.2028 , epoch: 221
E_val: 0.2122 , epoch: 222
E_val: 0.2114 , epoch: 223
E_val: 0.2058 , epoch: 224
E_val: 0.3746 , epoch: 225
```

左邊為有做 feature normalization 所跑出的 Error of validation set，右邊則為沒有做的結果，所使用的 Error function 為 0/1 error，可看到有做特徵標準化的效果確實較好，也有較為正常的訓練趨勢，沒有進行特徵標準化的結果則會有跳躍式的結果，不僅難以取捨參數的抉擇，跑出的結果也較差，從下圖可以看到 kaggle 的結果上面為有做，下面則為沒做的結果，兩者其餘參數皆為固定，對資料分佈較為廣的 feature 進行 normalization 對於模型的準確度有很好的表現。

submission.csv a day ago by r06631035_王凱陞 For 第四題，有做feature normalization	0.81120	0.81180
submission.csv a day ago by r06631035_王凱陞 FOR 第四題 沒做feature normalization	0.57860	0.57020

5. 根據題意已知:

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

欲證明$\int_{-\infty}^{\infty} f(x)\, dx = 1$

Let $\frac{x-\mu}{\sigma} = u, \frac{dx}{\sigma} = du \to dx = \sigma du$

$\to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$

Let $I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$

$\to I^2 = (\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du)(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv)$

$= \frac{1}{2\pi} (\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(u+v)^2}{2}} du dv)$

Let $u = r\cos\theta, v = r\sin\theta, 0 \le \theta \le 2\pi, 0 \le r \le \infty$

$\to \frac{1}{2\pi} (\int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} r dr d\theta)$

Let $\omega = r^2, d\omega = 2rdr \to rdr = \frac{1}{2} d\omega$

$\to \frac{1}{2\pi}\frac{1}{2} (\int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{\omega}{2}} d\omega d\theta)$

$= \frac{1}{2\pi}\frac{1}{2} (\int_{0}^{2\pi} -2(e^{-\frac{\omega}{2}}) |_{\omega=0}^{\omega=\infty} d\theta)$

$= \frac{1}{2\pi} \left(\int_{0}^{2\pi} -(0-1)\, d\theta\right) = \frac{1}{2\pi}(2\pi) = 1$

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

故得證。

6.

(a) $\frac{\partial \mathrm{E}}{\partial z_k}$，Error function 會和$y_k$有關，根據題意$y_k = \mathrm{g}(z_k)$，故可展開如下:

$$\frac{\partial \mathrm{E}}{\partial z_k} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial z_k}$$

(b) $\frac{\partial \mathrm{E}}{\partial z_j}$，延續上題結果，又$z_k = \sum_j w_{jk} y_j$ ，$y_j = \mathrm{g}(z_j)$，故可展開為:

$$\frac{\partial \mathrm{E}}{\partial z_j} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial z_j}$$

(c) $\frac{\partial \mathrm{E}}{\partial w_{ij}}$，延續上題結果，又$z_j = \sum_j w_{ij} y_i$，故可展開為:

$$\frac{\partial \mathrm{E}}{\partial w_{ij}} = \frac{\partial \mathrm{E}}{\partial y_k}\frac{\partial y_k}{\partial z_k}\frac{\partial z_k}{\partial y_j}\frac{\partial y_j}{\partial z_j}\frac{\partial z_j}{\partial w_{ij}}$$

<探討> 此題雖然只需做基本的偏微分展開，但由最後的 Error function 從最後一層推導回來，即是 backpropagation 的做法。