# Homework 2 Report Problem Set

Professor Pei-Yuan Wu

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，並試著討論可能原因。

| | Public score | Private score |
|---|---|---|
| logistic regression | 0.81220 | 0.80700 |
| generative model | 0.81200 | 0.80620 |

由上表可知logistic regression的表現略勝於generative model，我認為這是因為generative model考慮到一些非實際的情況而影響它的準確度。

generative model是假設於model data是來自於一個機率模型，也就是說它會把model的所有可能情況考慮進模型裡。它會去考慮input feature的所有可能情況，即使該種情況實際上並沒有在training data及test data中，或者該種情況現實上沒可能出現，generative model也會把這些可能性考慮進去，導致它會考慮一些data以外的情況，影響最後得到的model，最終導致model的準確率下降。

**Problem 2.** (1%) 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。

| | Public score | Private score |
|---|---|---|
| No one-hot encoding | 0.80600 | 0.79720 |
| one-hot encoding | 0.81160 | 0.80660 |

在原來的數據集中，是透過把gender, education, marital status這幾種特徵數據轉化為有序的數字序列來表示，以方便進行運用及訓練。但是，即使轉化為數字表示後，上述特徵也不能直接用在我們的分類器中。因為，分類器會認為特徵數據是連續的，並且是有序的。

而事實上這些特徵是從透過轉換而得來，並不存在連續性及有序性。經過one-hot encoding後，如果一個特徵有n種不同的值，它會變成了n個二元特徵。並且，這些特徵互斥，因此，數據會變成稀疏的。

所以one-hot encoding替我們解決了分類器不好處理屬性數據的問題，令模型準確率有所提升。

**Problem 3.** (1%) 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）。

| Delete features | Public score | Private score |
|---|---|---|
| **None** | **0.81200** | **0.80620** |

| (train by ALL features) | | |
|---|---|---|
| LIMIT_BAL | 0.81180 | 0.80540 |
| SEX | 0.81140 | 0.80860 |
| EDUCATION | 0.81140 | 0.80740 |
| MARRIAGE | 0.81140 | 0.80800 |
| AGE | 0.81300 | 0.80780 |
| PAY_0 | 0.80000 | 0.79840 |
| PAY_2 | 0.80980 | 0.80980 |
| PAY_3 | 0.81300 | 0.80640 |
| PAY_4 | 0.81200 | 0.80600 |
| PAY_5 | 0.81140 | 0.80620 |
| PAY_6 | 0.81180 | 0.80620 |
| BILL_AMT1 | 0.81120 | 0.80600 |
| BILL_AMT2 | 0.81200 | 0.80620 |
| BILL_AMT3 | 0.81160 | 0.80600 |
| BILL_AMT4 | 0.81200 | 0.80600 |
| BILL_AMT5 | 0.81180 | 0.80620 |
| BILL_AMT6 | 0.81200 | 0.80620 |
| PAY_AMT1 | 0.81160 | 0.80560 |
| PAY_AMT2 | 0.81200 | 0.80540 |
| PAY_AMT3 | 0.81180 | 0.80600 |
| PAY_AMT4 | 0.81180 | 0.80640 |
| PAY_AMT5 | 0.81180 | 0.80580 |
| PAY_AMT6 | 0.81200 | 0.80580 |

首先，我是透過generative model進行訓練，並於每次訓練時各刪去一個input features來進行比較，比較基準為若Public score及Private score均比使用所有features訓練的分數都要低，則認為它是影響較大的features。

即是當Public score/Private score同時低於0.81200/0.80620時，可考慮成該features的重要性比較高，上表則是所有features的比較結果，紅字標示的部份為該features被刪去後訓練所得的分數比利用所有features訓練所得的分數都要低。

但是，考慮到” PAY_0,2~6” 、” BILL_AMT_1~6” 及” PAY_AMT_1~6” 這幾組features存在著與時間上的連續性，我修正了我的實驗方式如下。

| Delete features | Public score | Private score |
|---|---|---|
| **None (train by ALL features)** | **0.81200** | **0.80620** |
| LIMIT_BAL | 0.81180 | 0.80540 |
| SEX | 0.81140 | 0.80860 |
| EDUCATION | 0.81140 | 0.80740 |
| MARRIAGE | 0.81140 | 0.80800 |
| AGE | 0.81300 | 0.80780 |
| PAY_0,2-6 | 0.78060 | 0.78120 |
| BILL_AMT1-6 | 0.80920 | 0.80760 |
| PAY_AMT1-6 | 0.81080 | 0.80580 |

我將我的實驗方式修正成在刪去features時，把” PAY_0,2~6” 、” BILL_AMT_1~6” 及” PAY_AMT_1~6” 各視為一組features，一併刪去來進行測試。

結果如上表所示，可看出” LIMIT_BAL” 、” PAY_0,2~6” 、” BILL_AMT_1~6” 及” PAY_AMT_1~6” 這4組features在刪去後所得到的分數都有下降，可認為它們的影響較大，當中又以” PAY_0,2-6” 下降的幅度最大，影響也應是最大。

**Problem 4.** (1%) 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。

| | Public score | Private score |
|---|---|---|

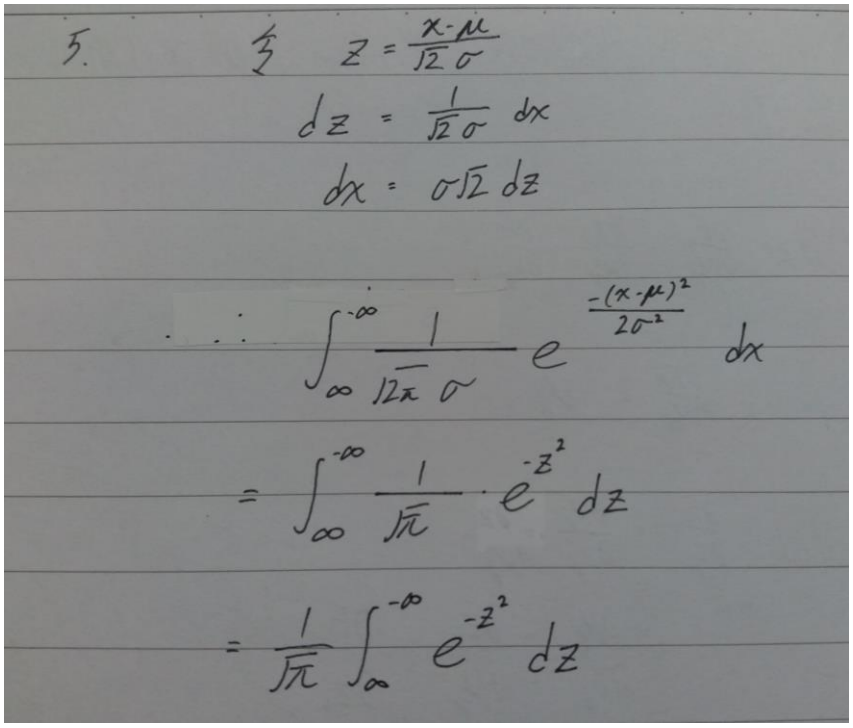
| | | |
|---|---|---|
| No normalization | 0.79960 | 0.79460 |
| feature normalization | 0.81180 | 0.80800 |

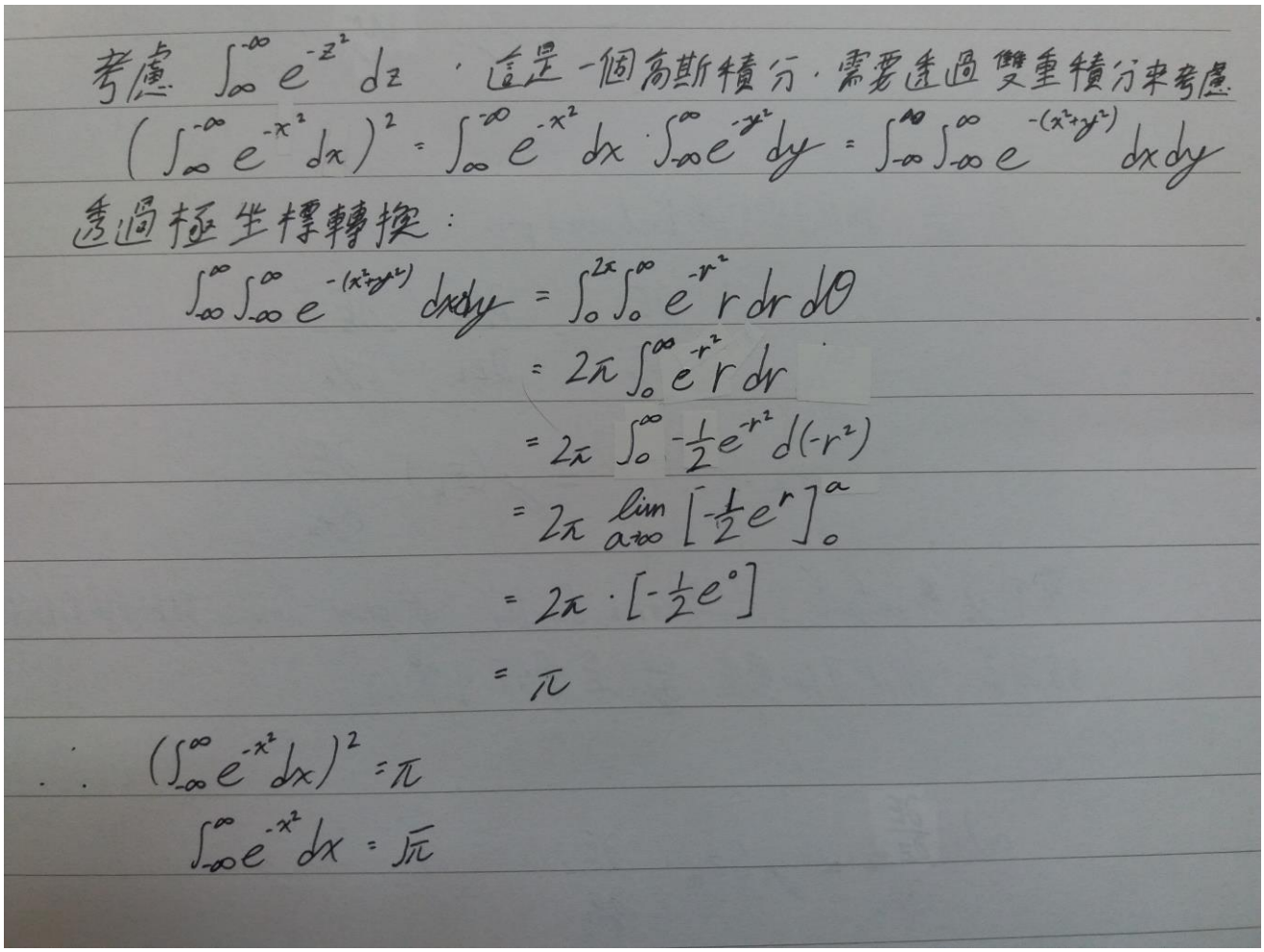
在資料集中擁有各種的feature，當中不同的feature的可能值的範圍可以很大也可以很小，像”PAY_1~6“及”BILL_AMT1~6” ，”PAY_1~6“的可能值的範圍較小，”BILL_AMT1~6” 的可能值範圍比較大。

當每個feature乘上各自的weight時，數值比較大的feature對結果的影響明顯會比較大，而因為Model背後是用空間中的距離來做區分，假設某一個特徵過大，該Model的成本函數會被這個特徵所支配。

所以為了避免某些feature的數值太大而影響了model的訓練結果，我們需要把feature的值限定在某個範圍內，在這邊我用了min-max normalization，將feature限定在0~1之間，成功令model準確率有所提高。

**Problem 5.** (1%)


5. 令 $z=\frac{x-\mu}{\sqrt{2}\sigma}$
$dz=\frac{1}{\sqrt{2}\sigma}dx$
$dx=\sigma\sqrt{2}\,dz$
$\therefore \int_{\infty}^{-\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx$
$=\int_{\infty}^{-\infty}\frac{1}{\sqrt{\pi}}\cdot e^{-z^2}dz$
$=\frac{1}{\sqrt{\pi}}\int_{\infty}^{-\infty}e^{-z^2}dz$


考慮 $\int_{-\infty}^{\infty}e^{-z^2}dz$，這是一個高斯積分，需要透過雙重積分來考慮
$\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)^2=\int_{-\infty}^{\infty}e^{-x^2}dx\int_{-\infty}^{\infty}e^{-y^2}dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(x^2+y^2)}dx\,dy$
透過極坐標轉換：
$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(x^2+y^2)}dx\,dy=\int_{0}^{2\pi}\int_{0}^{\infty}e^{-r^2}r\,dr\,d\theta$
$=2\pi\int_{0}^{\infty}e^{-r^2}r\,dr$
$=2\pi\int_{0}^{\infty}-\frac{1}{2}e^{-r^2}d(-r^2)$
$=2\pi\lim_{a\to\infty}\left[-\frac{1}{2}e^{r}\right]_0^a$
$=2\pi\cdot\left[-\frac{1}{2}e^{0}\right]$
$=\pi$
$\therefore\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)^2=\pi$
$\int_{-\infty}^{\infty}e^{-x^2}dx=\sqrt{\pi}$

$$\therefore \left(\int_{-\infty}^{\infty} e^{-x^2} dx\right)^2 = \pi$$

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

$$\Rightarrow \frac{1}{\sqrt{\pi}} \cdot \int_{-\infty}^{-\infty} e^{-z} dz = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1$$

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

**Problem 6.** (1%)

6. 對於任意的 $\frac{\partial E}{\partial w}$，我們可以利用 chain rule 進行拆分
得出 $\frac{\partial E}{\partial w} = \frac{\partial z}{\partial w} \cdot \frac{\partial E}{\partial z}$

所以，我們得知 $\frac{\partial E}{\partial w_{ij}} = \frac{\partial z_j}{\partial w_{ij}} \cdot \frac{\partial E}{\partial z_j}$

利用 Forward pass：$\frac{\partial z_j}{\partial w_{ij}} = y_i$

利用 Backward pass：$\frac{\partial E}{\partial z_j} = \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial E}{\partial y_j}$

$$= \frac{\partial y_j}{\partial z_j} \cdot \frac{\partial z_k}{\partial y_j} \cdot \frac{\partial E}{\partial z_k}$$

$$= g'(z_j) \cdot w_{ij} \cdot \frac{\partial E}{\partial z_k}$$

對 $\frac{\partial E}{\partial z_k}$，我們再利用 Backward pass：

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

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

當中 $y_k$ 為 output，E 為 error function，而 error function 對於不同的方法都有著它的明確定義，所以可以考慮 $\frac{\partial E}{\partial y_k}$ 是明確可算的

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

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

c). $\frac{\partial E}{\partial w_{ij}} = y_i \cdot \frac{\partial E}{\partial z_j} = y_i \cdot g'(z_j) \cdot w_{jk} \cdot g'(z_k) \cdot \frac{\partial E}{\partial y_k}$