Machine Learning HW2 Report

B05901043 電機三 莊鎧爾

Problem 1

Use the Kaggle public leader board score for discussion.

Model Type	Score on Kaggle set
Generative model	0.81200
Logistic regression model 1	0.81040
Logistic regression model 2	0.81240

Logistic regression model 1 uses all dimensions of training data and logistic regression model 2 uses only 7 dimensions of training data. In the generative model, when I use different covariance matrix in two gaussian distribution, I will get a worse result. But when I use the same covariance matrix, the score will become higher. It is the same in the logistic regression model. When I use full date in the model, the model tends to overfitting. Therefore, reducing the data dimension make the model become better. Besides, the logistic regression model 2 not yet implement feature normalization, but the result is already better than the generative model. It is because the generative model is limited by the gaussian distribution model, the model cannot create much better weights to fit the training data.

Problem 2

I divide the data into two parts, 15000 data as training set and 5000 data as validation set

Model Type	Correct rate in validation set
With one-hot	0.8110
No one hot	0.8104

When the gender, education, and marital status use one-hot encoding, the result will become better, but it just improves a little. Because the numbers recording inside these data cannot represent the level of such data. For example, the relationship between marital status 2 and marital status 1 is not closer than the relationship between status 4 and status 1. Therefore, the data cannot be record as number, and in replacement, we use the one-hot encoding.

Problem 3

I try to use just one dimension on model training, but I find that there are many dimensions producing a bad model. Th model guesses 0 for all testing data. Therefore, I try another method. The procedures I have done are as following:

1. Just using one dimension to train a model, after all dimensions are trained, I choose the best model and add that dimension to a dimension list
2. Using dimensions inside the dimension list and another dimension to train model, choose the best model and add that dimension to the dimension list.
3. Repeating this procedure until all dimensions are added into the dimension list.

By the procedures, I know that if a feature is more important, then it will be added into the dimension list earlier. The order of dimensions inside the dimension list is as following:

$$[5, 6, 9, 7, 10, 8, 0, 1, 2, 3, 4, 11, 18, 21, 17, 12, 15, 13, 19, 20, 14, 16, 22]$$

Therefore, we know Pay0 to Pay6 is the important features in this classification testing, and BILL_AMT and PAY_AMT do not improve the model much.

Problem 4

I divide the data into two parts, 15000 data as training set and 5000 data as validation set

Model Type	Correct rate in validation set
With normalization	0.8136
No normalization	0.7306

I find that if the data is not normalized before training, the model cannot easily train, because when I call the sigmoid function, the value often overflows. Therefore, I have to set the training rate as a small value, and this make the model weights update slowly. In the figure, we can see the model with no normalization producing a bad result. On the other hand, the normalization data can create a better model, so it brings a good result in the validation set.

Problem 5

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \Phi(\infty) - \Phi(-\infty) = 1 - 0 = 1$$

Problem 6

(a)

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

(b)

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

(c)

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial z_j} \times \frac{\partial z_j}{\partial w_{ij}} = \frac{\partial E}{\partial y_k} \times g'(z_k) \times w_{jk} \times g'\left(z_j\right) \times z_i$$