Name: Huiting Hong
ID: A061610
13 April 2018

Deep Learning Assignment 1

— Regression and Classification

Problem1(Regression):

Goal: Analyzing the energy efficiency of buildings by using the dataset which contains the simulation energy loads of different buildings using 8 different features.

Data: Shown in below

	0	1	2	3	4	5	6	7	8	9
0	# Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	Cooling Load
1	0.98	514.50	294.00	110.25	7.00	2.00	0.00	0.00	15.55	21.33
2	0.98	514.50	294.00	110.25	7.00	4.00	0.00	0.00	15.55	21.33
3	0.98	514.50	294.00	110.25	7.00	5.00	0.00	0.00	15.55	21.33
4	0.90	563.50	318.50	122.50	7.00	2.00	0.00	0.00	20.84	28.28
5	0.90	563.50	318.50	122.50	7.00	3.00	0.00	0.00	21.46	25.38
6	0.90	563.50	318.50	122.50	7.00	4.00	0.00	0.00	20.71	25.16
7	0.86	588.00	294.00	147.00	7.00	2.00	0.00	0.00	19.50	27.30
8	0.86	588.00	294.00	147.00	7.00	3.00	0.00	0.00	19.95	21.97
9	0.86	588.00	294.00	147.00	7.00	4.00	0.00	0.00	19.34	23.49
10	0.82	612.50	318.50	147.00	7.00	3.00	0.00	0.00	17.41	21.46
11	0.82	612.50	318.50	147.00	7.00	4.00	0.00	0.00	16.95	21.16

- 8 features (6 numerical + 2 categorical)
- 769# in Total

Target: Heating Load

(i) Construct DNN → Apply Feature Selection

- Construct DNN :
  - Structure:
    - **15** - ReLU - **10** - ReLU - **10** - ReLU - **1** (Remove 1st/2nd/3rd/4th/5th/7th feature)
    - or **12** - ReLU - **10** - ReLU - **10** - ReLU - **1** (Remove 6th feature)
    - or **10** - ReLU - **10** - ReLU - **10** - ReLU - **1** (Remove 8th feature)

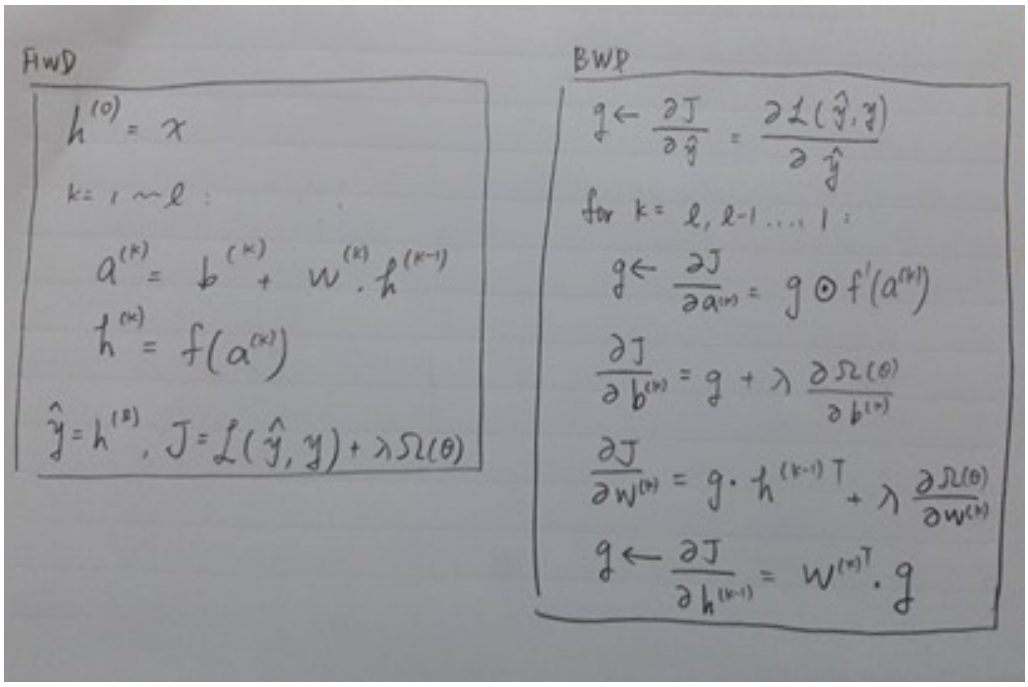
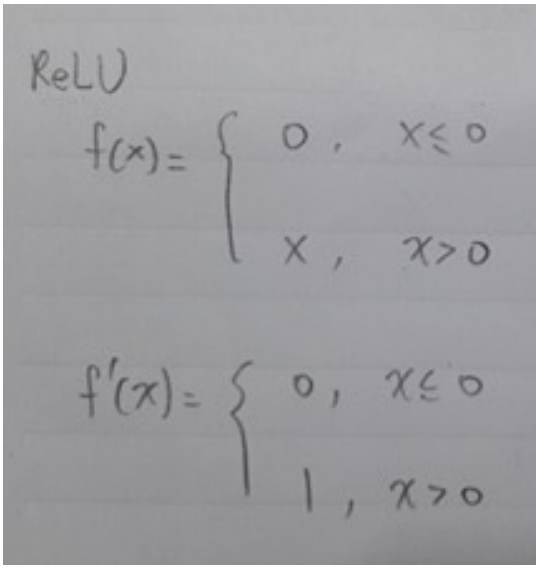


Figure1: The formula of forward propagation(on the left hand side) and the backwardpropagation(on the right hand side).

- Algorithm : Using Forward Propagation & Backward Propagation (Formula is shown in Figure1), and the error function is defined as square loss. We finally measure the performance using Root Mean Square Error.
- Note: My activation function is using ReLU, so we need both its formula and its derivative function :



- Result:

- The network architecture, the features we choose, the training RMS error and the test RMS error are all shown in the Table 1.
- The leaning curve, regression result with training/testing labels are shown in Figure 2~9.

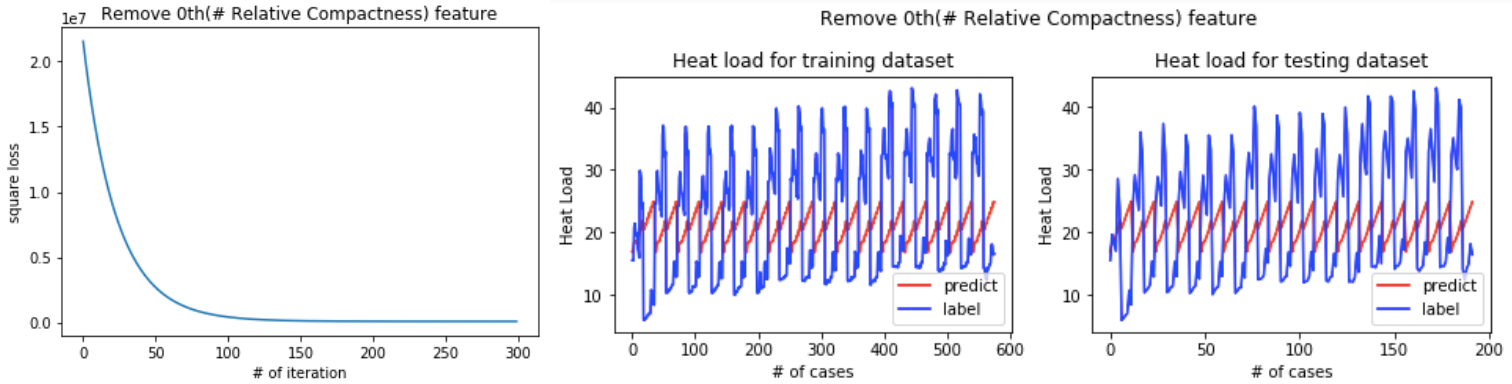


Figure2: Remove 1st feature, # of Relative Compactness, we can see the learning curve decrease in a good shape, however the regression result is not as good as expect. Although there did exist the sign which predict result are trying to be close to the ground truth.

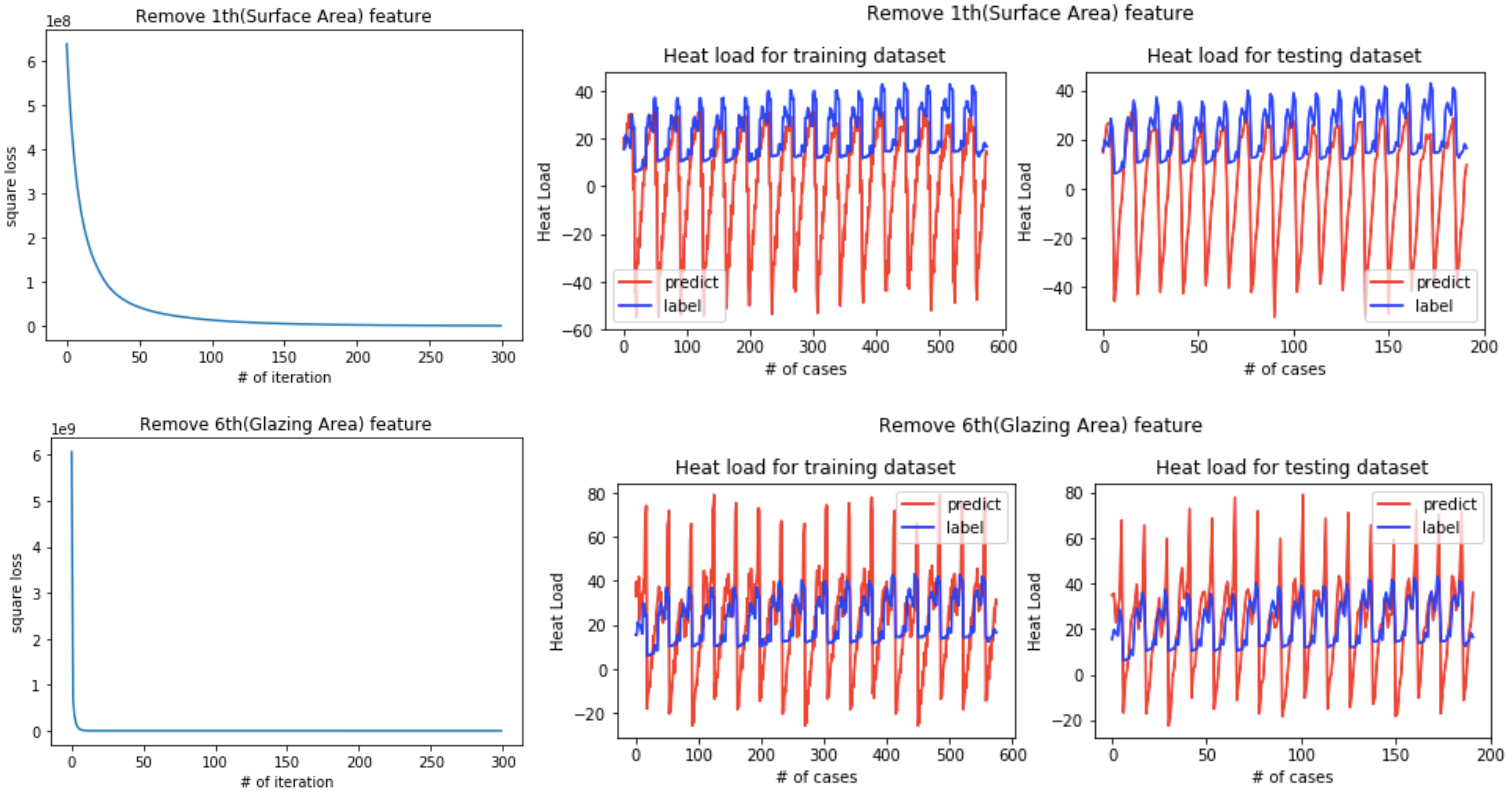


Figure3: Remove 2nd feature & Remove 7th feature respectively. We can see the learning curve decrease more abruptly, and the predict result fit the ground truth closer than the result in Figure2.

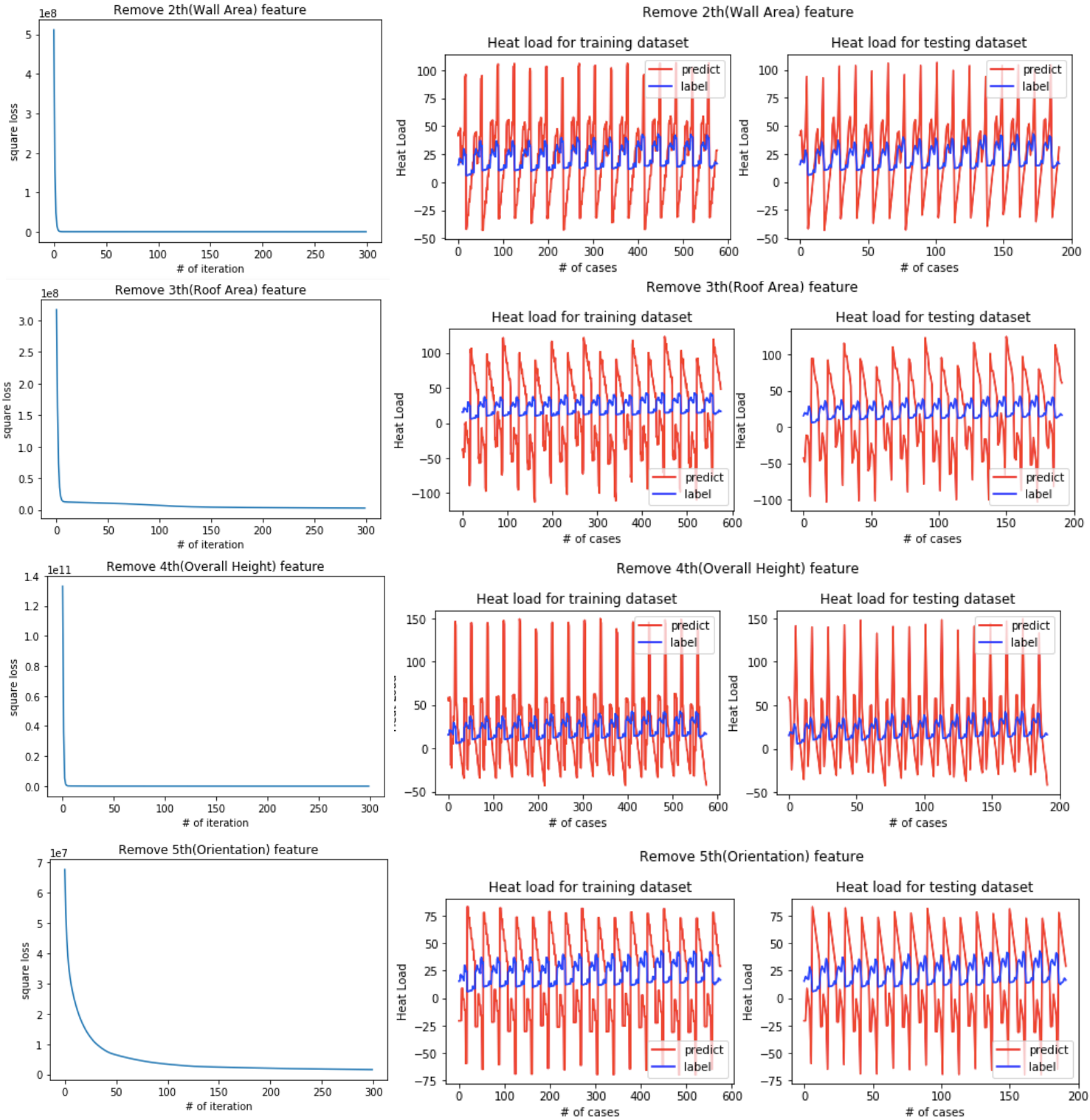


Figure4: Remove 3rd/ 4th/ 5th/ 6th feature respectively. We can see the learning curve decrease pretty steep, and the predict result is not as good as expected.

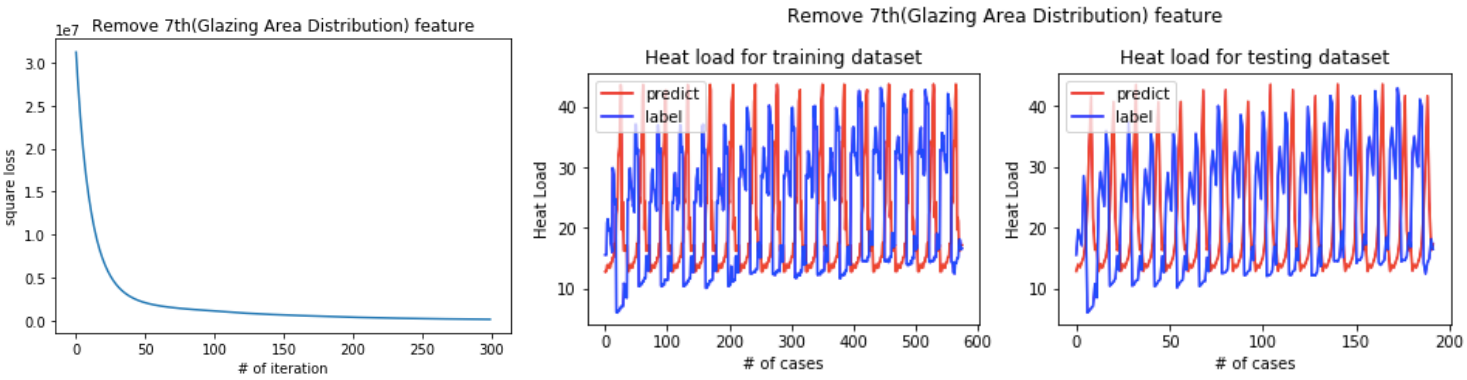


Figure4: Here the result fit SUPER GOOD! Therefore I think remove the 7th feature is the best choice, and its learning curve is also decreasing in a smooth way.

Table 1

Network Archtecture	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	15-10-10-1	12-10-10-1	15-10-10-1	10-10-10-1
Selected Features	[1,2,3…7]	[0,2,3…7]	[0,1,3…7]	[0,1,2,4…7]	[0..3,5…7]	[0…4,6,7]	[0…4,5,7]	[0…4,5,6]
Training RMS	**11.54**	27.52	28.67	69.37	41.80	52.43	**16.24**	**16.92**
Testing RMS	**11.54**	27.23	28.61	69.33	41.59	52.36	**15.93**	**16.90**

(ii) Discuss which input features significantly affect the heating load

base on the learning curve, RMS error and the regression result, I think the 1st&6th&7th are the most important features that will affect the heating load. Which is # Relative Compactness, Glazing Area and Glazing Area Distribution.

Problem2(Classification):

Goal: Base on the attributes obtained from the emails to identify spam on email automatically.

(i) Mine variables

- number of hidden layers: 4

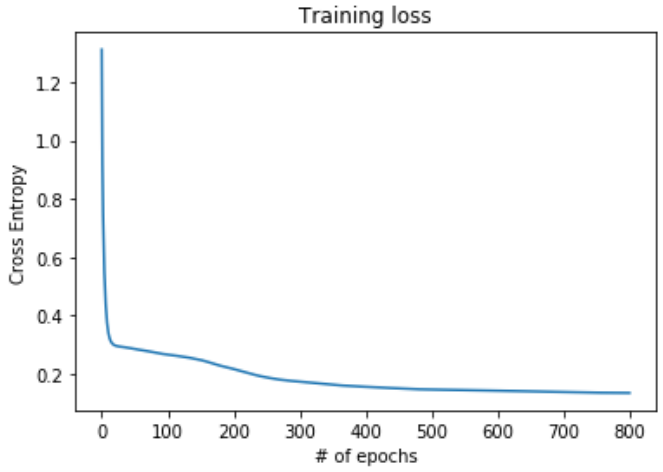
- number of hidden units: 40 → 10 → 3 or 2 → 2

- learning rate: 10e-5

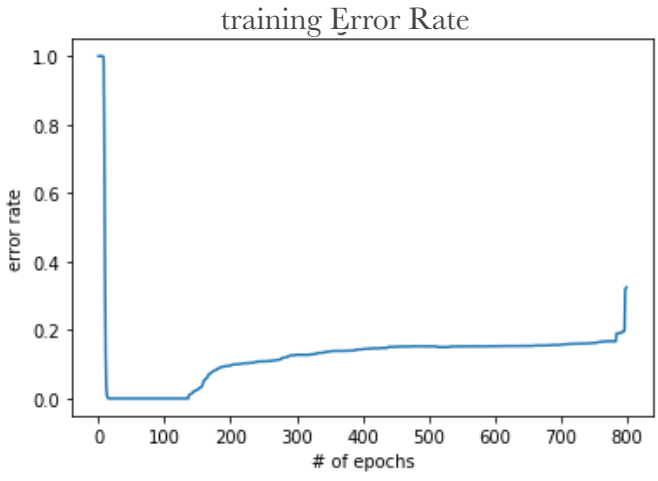
- number of iterations: 800

- mini-batch size : I didn't apply mini-batch approach, so here I set to 2000, which is exactly the data number.

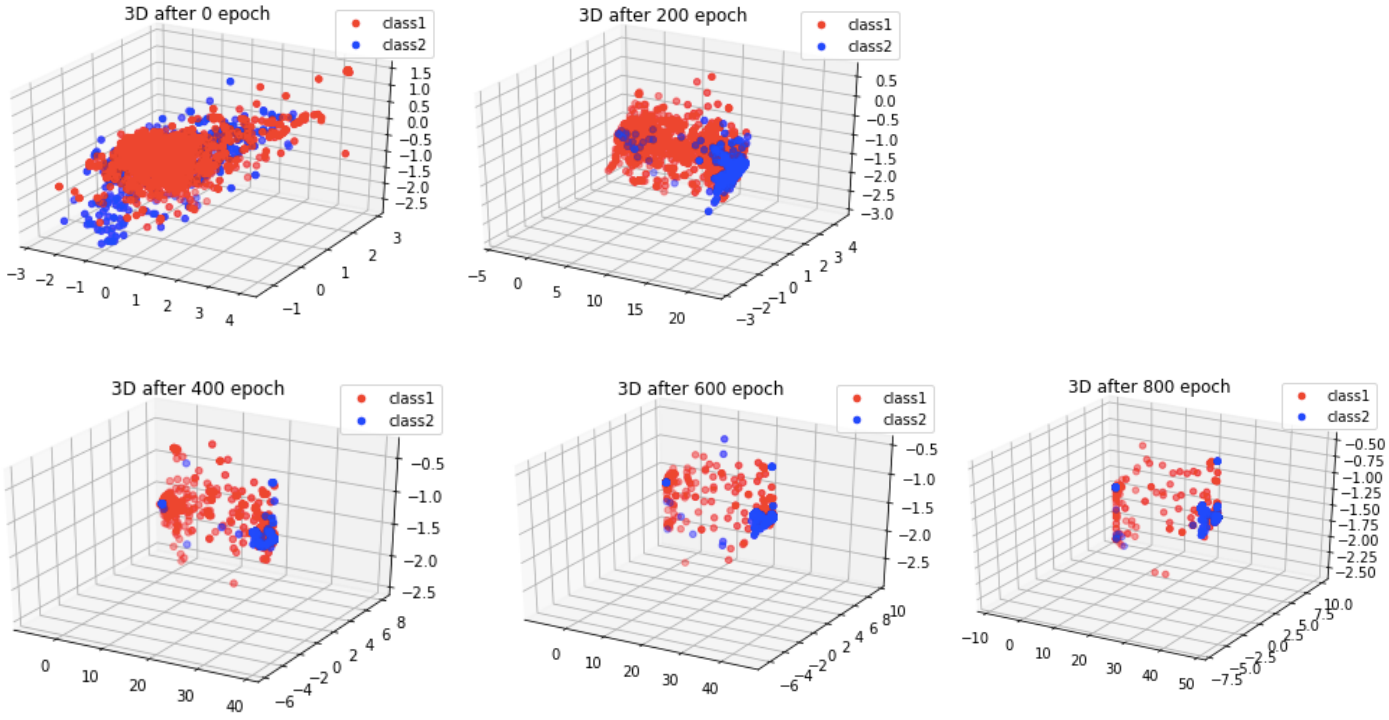
(a) Learning Curve

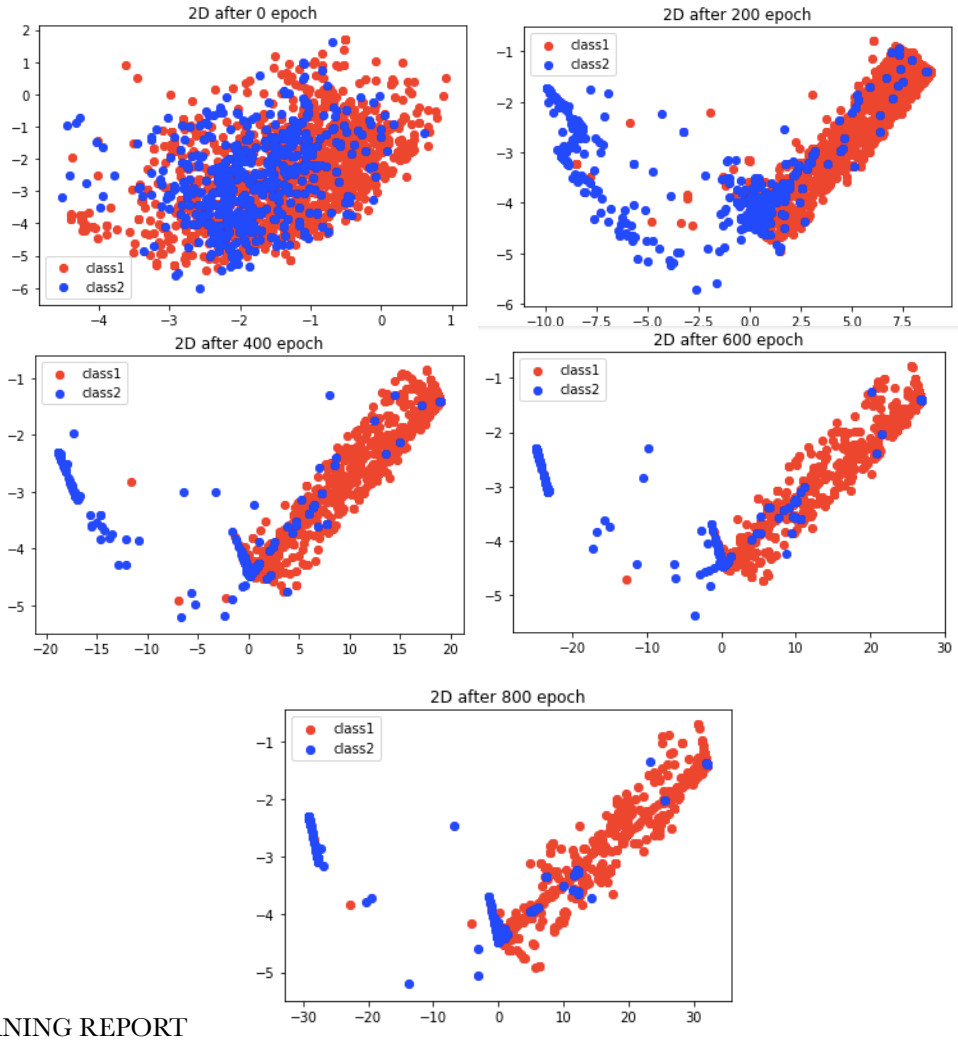


(b) training/testing error rate



(ii) Distributions of latent features (I did Both 3 node and 2 node)





As we can see when the iteration increases, the cluster become more clear, which means that the classification result performs better when iteration in creases.