Homework1 – Regression & Classification Models

林峻賢 312581018

1. **REGRESSION**

a) Please try to predict the heating load of buildings by minimizing the sum-of-squares error function.

b) Show your (1) <mark>network architecture</mark> (number of hidden layers and neurons), (2) <mark>learning curve</mark>, (3) <mark>training RMS error</mark>, (4) <mark>test RMS error</mark>, (5) <mark>regression result with training labels</mark> and (6) <mark>regression result with test labels</mark> in the report.

I. <mark>Network architecture, training RMS error, test RMS error</mark>

Network Architecture	15-10-10-1
Training E-RMS	0.3343
Testing E_RMS	0.5222
Epochs	10000
Learning rate	0.0002
Batch size	128
Training size – Testing size	75% - 25%

Table I, Related regression model hyper parameters.

II. <mark>Learning curve</mark>
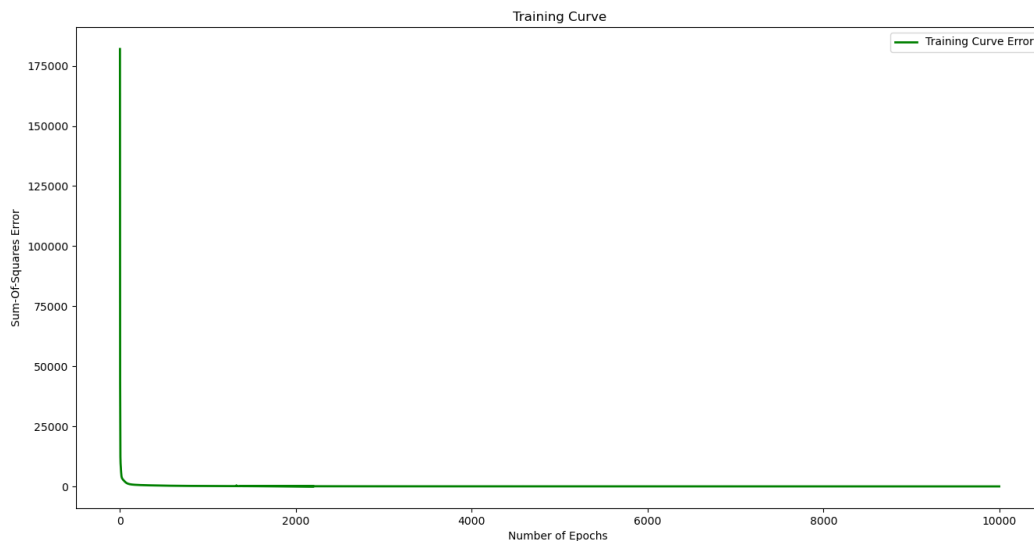

Figure I.   sum-of-squares error - # of epoch of training curve.

Figure II.   Regression results with training label. Green one is ground truth, and red one is the results of the regression model. We can see that the prediction almost matches to the ground truth, which means **the model has already fit the training data.**
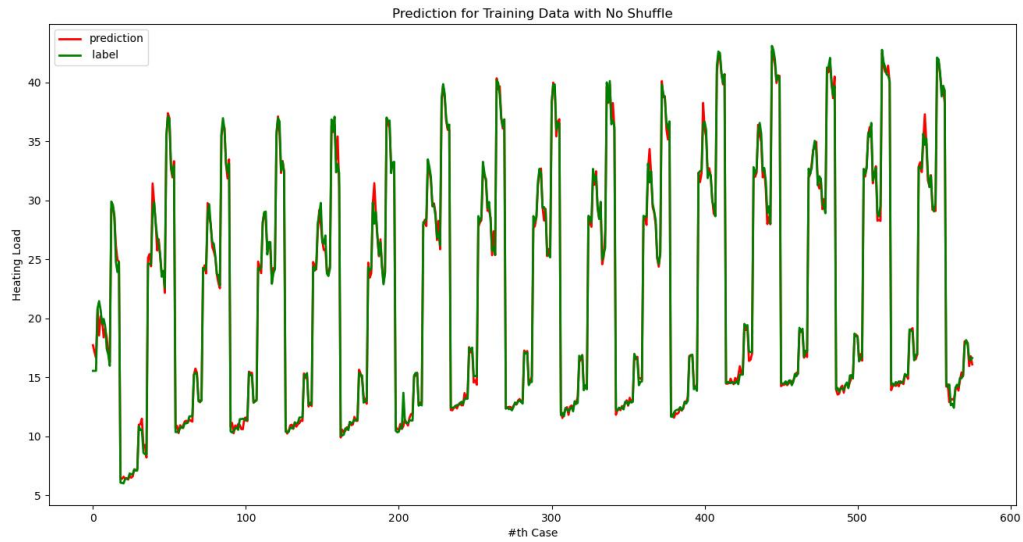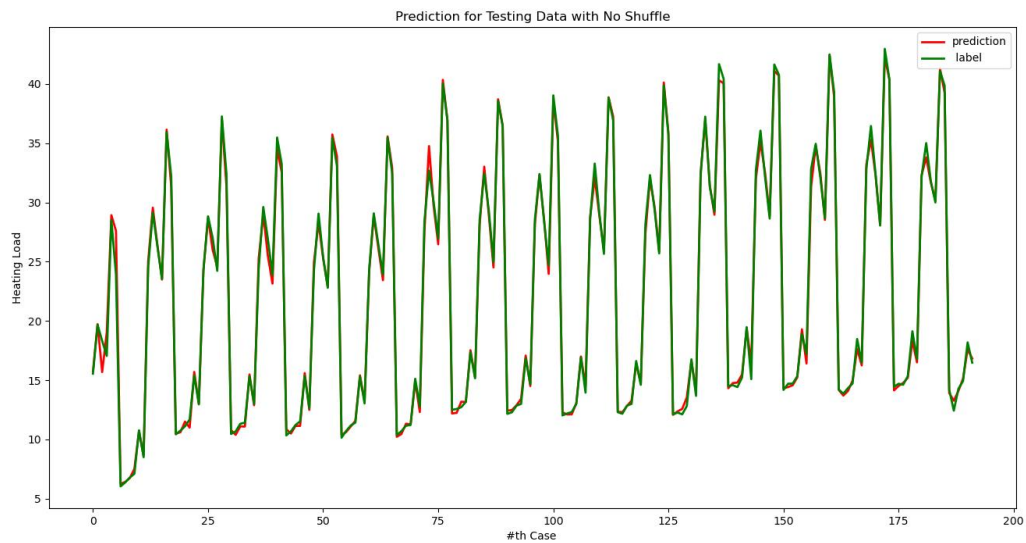
Figure III.   Regression results with test label. Green one is ground truth, and red one is the results of the regression model. We can see that the prediction almost match to the ground truth, which means **the model can properly predict the unseen data.**

c)	Design a feature selection procedure to find out which input features influence the energy load significantly and explain why it works. You may compare the performance of choosing different features.

I use regression for feature selection, **it calculates the correlation between a specific feature and the target column**. The larger the computed value, the more information that particular feature provides for making decisions related to the target column, so the feature should be retained. Conversely, when the correlation is smaller, it indicates that the feature contributes less information to the decision-making process and should be discarded. Based on this logic, we can obtain a ranking of the most important features for decision-making, allowing for a reduction in data volume to enhance computational efficiency.

The following displays the performance of features of different sizes on the model. I found that **when feature size is 10**, the model not only can get the better performance than the model using the whole features, but also reduce the computation amount.

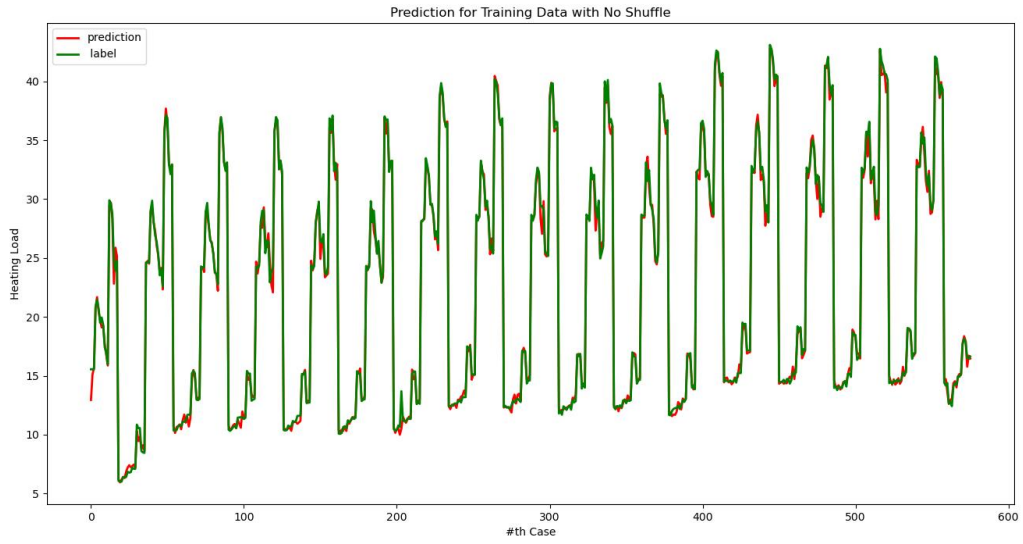I.   Select 15 features out of the original features.

Figure IV. Training prediction for 15 features, which has Root-mean-square error(RMS): **0.3952**
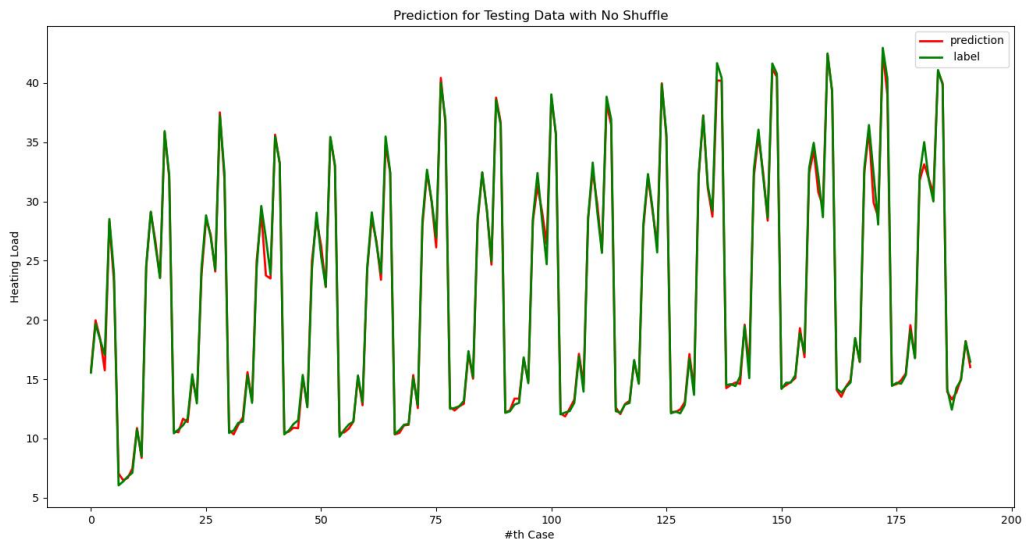

Figure V.   Testing prediction for 15 features, Root-mean-square error(RMS): **0.8132**

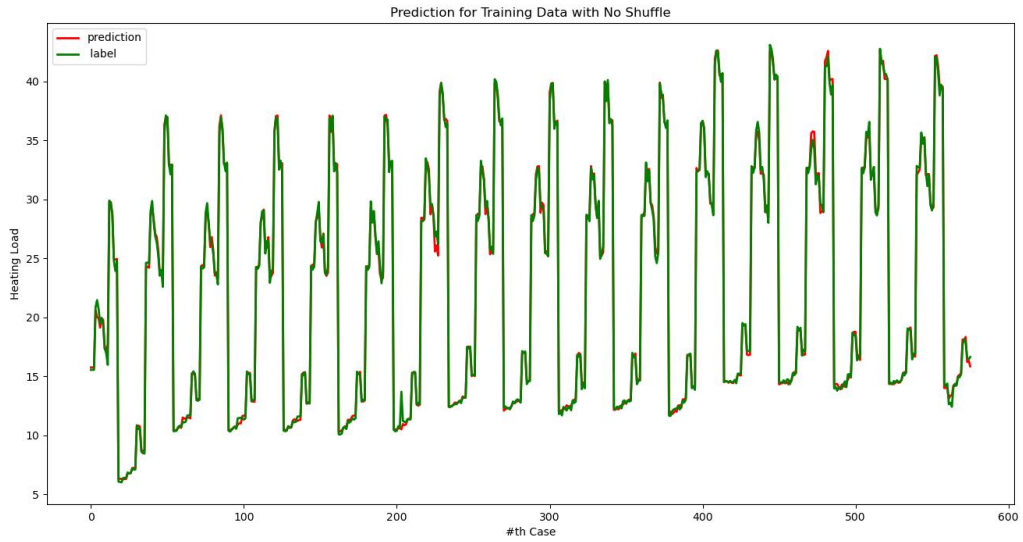II.   Select 10 features out of the original features.

Figure VI. Training prediction for 10 features, which has Root-mean-square error(RMS): **0.3410**
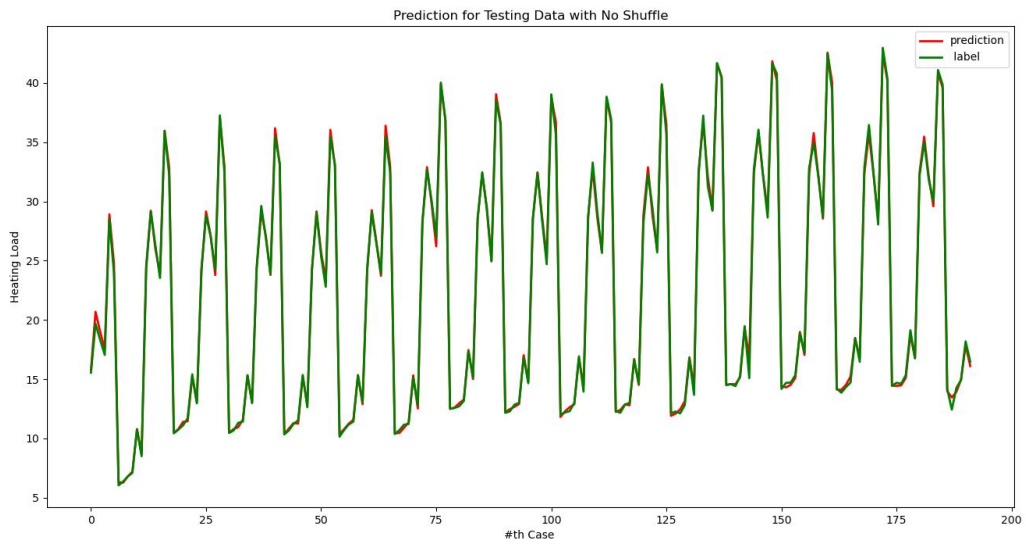

Figure VII. Testing prediction for 10 features, Root-mean-square error(RMS): **0.4793**

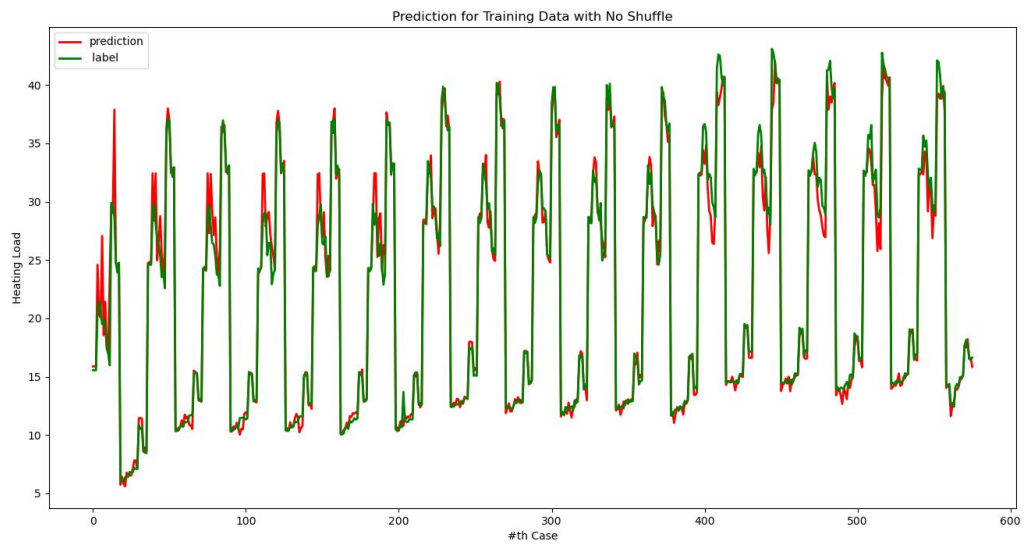III.   Select 5 features out of the original features.

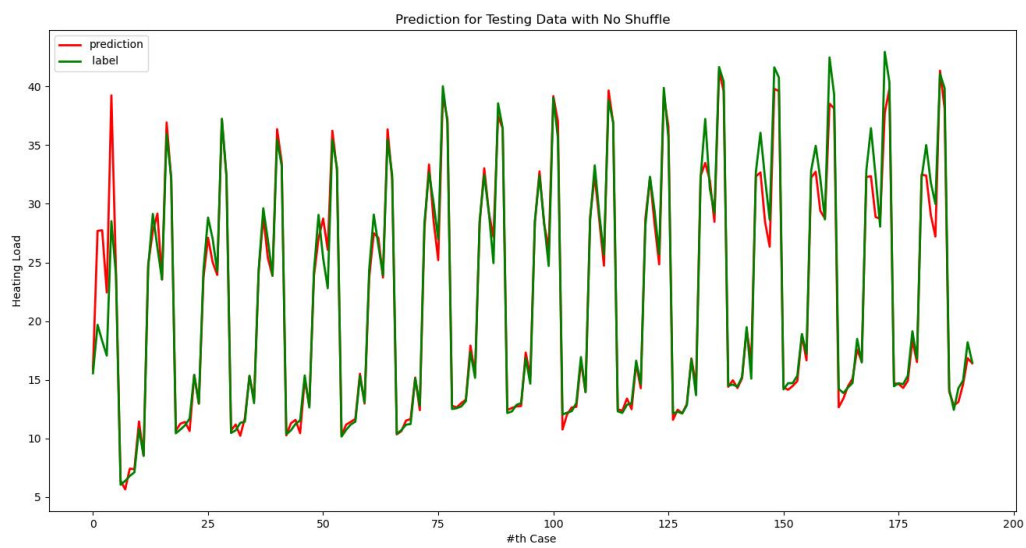Figure VIII.   Training prediction for 5 features, which has Root-mean-square error(RMS): **1.3989**

Figure IX.   Testing prediction for 5 features, Root-mean-square error(RMS): **1.6904**

2. CLASSIFICATION

a) Please try to classify the Ionosphere data by minimizing the cross-entropy error function.

b) Show your (1) network architecture (number of hidden layers and neurons), (2) learning curve, (3) training error rate, (4) test error rate in the report.

I. Network architecture, training RMS error, test RMS error

Network Architecture	33-16-2-1
Training error rate	0.0321
Testing error rate	0.1549
Epochs	50000
Learning rate	0.0002
Batch size	128
Training size – Testing size	80% - 20%

Table II, Related classification model hyper parameters.
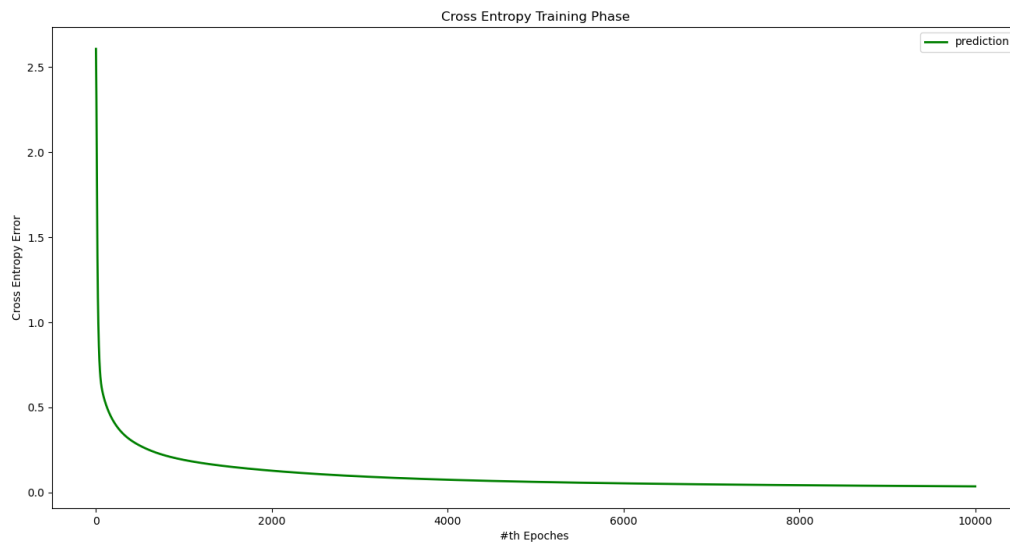
II. Learning curve
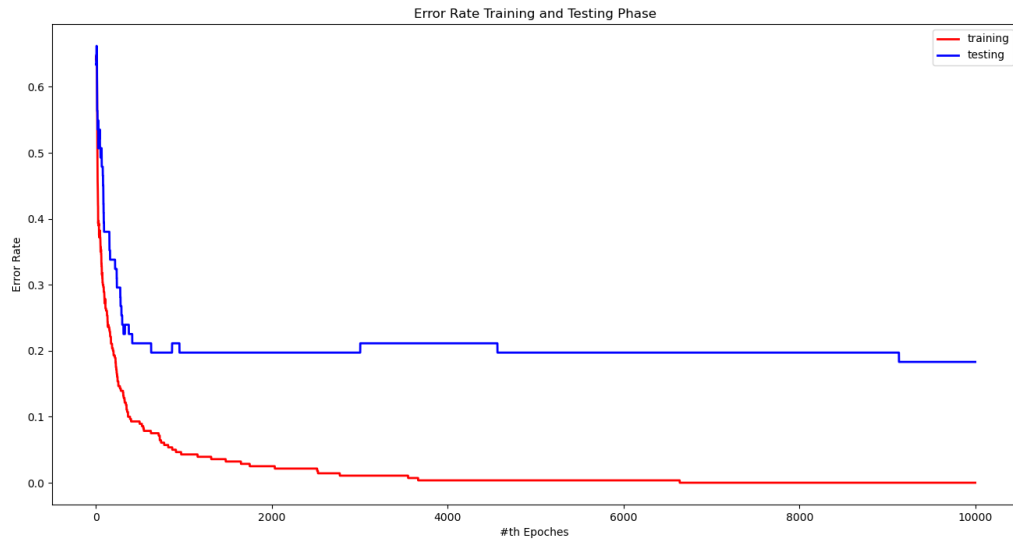

Figure X. Cross entropy at training phase.

Figure XI. Training error rate and test error rate

c)   Compare the results of choosing different numbers of nodes in the layer before the output layer by plotting the

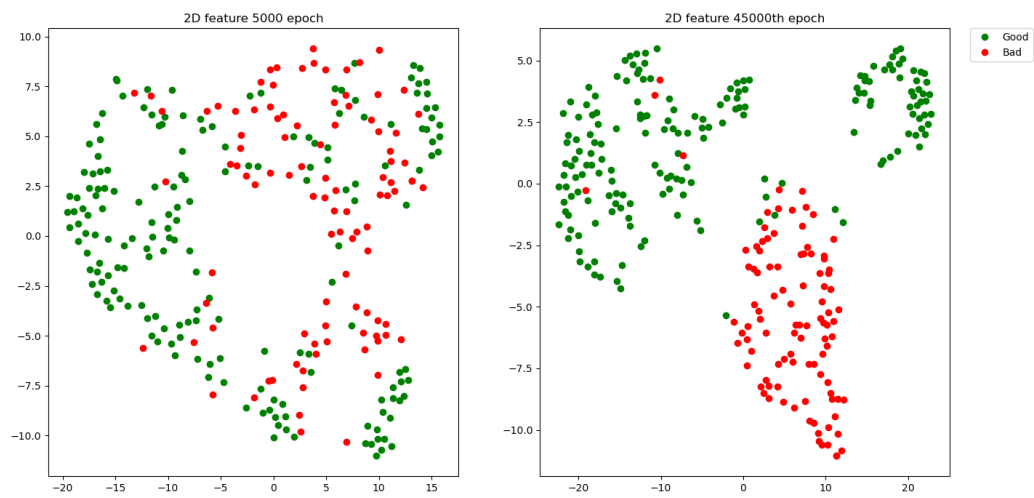distribution of latent features at different training stage.

Figure XII. Visualization the distribution of this hidden layer at diffident training stage.

By changing the hidden layer before output layer, using 2 nodes, then we can compare the result by visualize the distribution of this hidden layer at diffident training stage.