Project 1 report:

Part 0: import the toolkits that will be used in the project

From line7 to line13

To finish this project, I applied pytorch kit to assistant my tasks. Pytorch is used to establish machine learning model and train the model.

Part1: define classifier, build up neuron network

From line17 to line53

Firstly, define two gaussian sample cluster in 4 dimensions [line37 to line40], for each cluster, there are 10,000 data sets. With such data, I defined a 4-layers neuron network to obtain the output.

Part2: define the loss function

Form line57 to line61

The loss is based on Binary cross entropy, and the factor, Euclidian distance $\vec{x}$ and $\overrightarrow{\mu_1}$, is added to the end of loss function to form a new loss function.

Part3: network training

From line65 to line176

To begin with, create the dataset class, and combine the dataset to data loader [line65 to line95]. In data loader, I shuffled the data and separated them to train or to test. Then I composed the training part. For each epoch time, extracting input and target from data loader, I trained the model and stored the training history [line108 to line129]. Following is model evaluating, where I repeated use the data loader and store the result of evaluation [line131 to line145]. For ever 10 times, print the accuracy of the model. Out of the function, I set the parameters to initialize the data loader and control the training process by define the optimizer [line156 to line174]. Finally, save the history data into file.

Part4: test the model

From line181 to line197

The procedure is like what I did in training part that I extract the test data from data loader and evaluate the accuracy of the model. Then plot the frequency of data of different classes in bins.

Part5: calculate the separation

From line201 to line220

Extracting the evaluated outputs from Part4 and counting the numbers, of samples with prediction of their labels, located in each bins, I could compose 2 function to finish the task. In the function “separation_pro” [line201 to line208], I count the number. In the function “separation” [line210 to line218], the counts of both clusters were obtained and I accumulated the separation value.