Project 2 report

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

From line8 to line13

(pytorch in proj1)

Matplotlib.pyplot kit is used to plot histogram of the loss function result of model training and model evaluation.

Time kit is for recording the time of total training process and test process.

Part 1: define the unsupervised training model

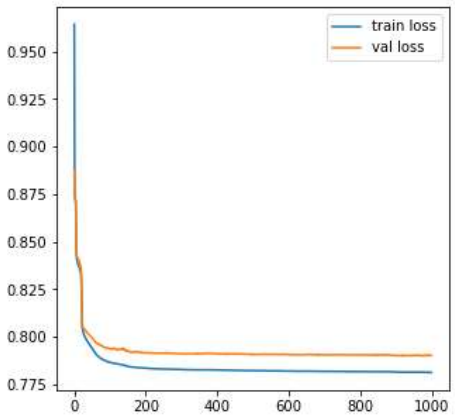
From line17 to line38

Compared with first project, Project 2 uses unsupervised training to train a model, then I should compose the autoencoder. In the autoencoder, I tested 2 phase setting and 3 phase setting. For 3 phases setting, the encoder matrix is [32*16], [16*8] and [8*4]. By comparing the results of 3 phases encoder and 2 phases encoder, I found that with more phases, the training period will shorter but with larger loss. With different requirement, the choices of masks are alternative.

Part 2: network training

From line40 to line147

To begin with, I composed the dataset and combined it into data loader [line40 to line61]. In the data loader, firstly, I shuffle the data of X_train and then cut it into 2 sets that one used for training while another used for evaluating. And then, I defined the training function [line64 to line112]. In the function, the model is an autoencoder type object, hence, for the model, unsupervised training is applied. Out of the function, I set the parameters to initialize the data loader and control the training. Finally, plotting the loss of model training procedure and saving the history data into file. The loss of training showed below.

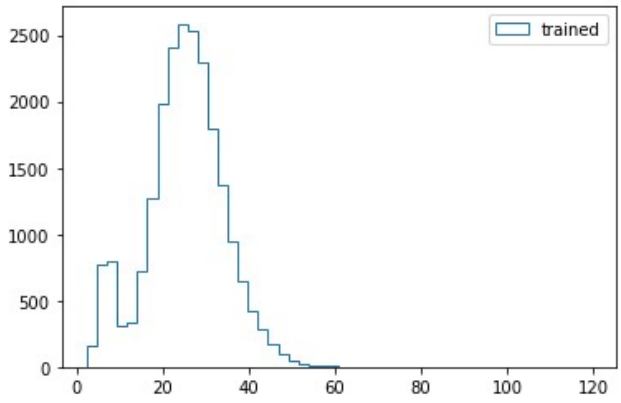


Part 3: test the model

From line149 to line154

The procedure is like the training, while the source of data is X_test. With loading the data and testing trained model with it, I could obtain the histogram of Mean Square

Error of the model. From the histogram, the valley of the MSE value is obvious. It is around 10. The histogram showed below.



Part 4: output the cut of MSE value

From line156 to line165

According to the histogram, the cut o the MSE is 1842.