Deep Learning Homework #6

Indra Imanuel Gunawan / 20195118

No. 3 – Anomaly Detection for Streaming Data using Autoencoder

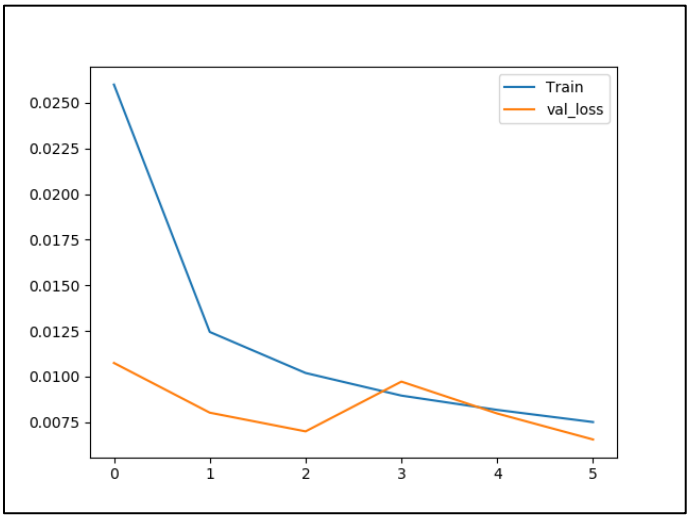
The code for this experiment can be found in my github repository "DL_Homework6", under the "No3" folder.

Similar with problem no 2, we are using autoencoder to detect anomalies in the dataset. The dataset that will be used in this experiment is the Forest dataset.

Note: This code is made using TensorFlow 1.4.0, running it using TensorFlow Version 2.x won't work.

There are 3 steps necessary to run this code:

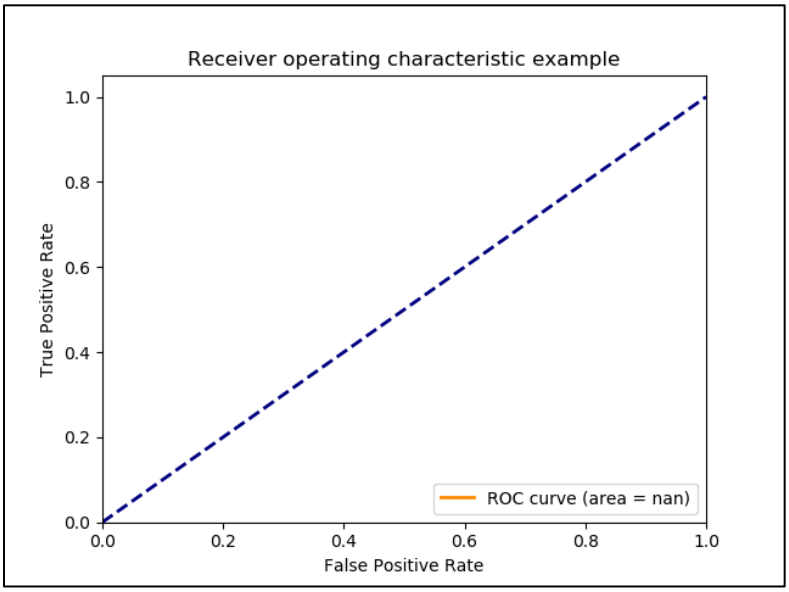
1. **Data Preparation**
   In this step, we are converting the raw data into uniform format using the dataPreparation.py. To run this code, type in "python src/DataPrepare/dataPreparation.py forest data/covtype.data.gz data" in the terminal. It'll create "FOREST.csv", the dataset that we'll use in this experiment.
2. **Initialization**
   In this step, we are initializing the model. To initialize the model, type in "python src/Initialization/initialization.py forest data/FOREST.csv outputs/model/" in the terminal. It'll create the model, and train it. Here are the training history for the model :




```
Epoch 0: Loss:0.026, Val_loss:0.011
Epoch 1: Loss:0.012, Val_loss:0.008
Epoch 2: Loss:0.010, Val_loss:0.007
Epoch 3: Loss:0.009, Val_loss:0.010
Epoch 4: Loss:0.008, Val_loss:0.008
Epoch 5: Loss:0.008, Val_loss:0.007
Got parameters mu and sigma.
254
Threshold:  0.008035432438127758
```

3. **Online Prediction**
   Finally, in this step, we're doing the online prediction, or the anomaly detection on the dataset. To run this script, type in "python src/OnlinePrediction/OnlinePrediction.py forest

data/FOREST.csv outputs/model/_8_25_10_para.ckpt.meta" in the terminal. It'll process 6000 batches of data and outputs a ROC (Receiver Operating Charateristic) graph:




```
5100 batches processed.
5200 batches processed.
5300 batches processed.
5400 batches processed.
5500 batches processed.
5600 batches processed.
5700 batches processed.
5800 batches processed.
5900 batches processed.
6000 batches processed.
```