

Music Classification based on Genre

Team Members: Aditi Laddha - 130050026
Palak Jain - 130050031
Dibyendu Mondal - 130050046

Problem Statement

The aim of the project is to develop a music genre classifier. There are many genres into which songs may be classified, e.g. pop, rock, classical, metal etc. Gigabytes of music files are spread over the web. To automate searching and organizing the music files based on their genre is a challenging task. In this report, we present our approach to classify genres of music file based on their content and analysis on some inputs.

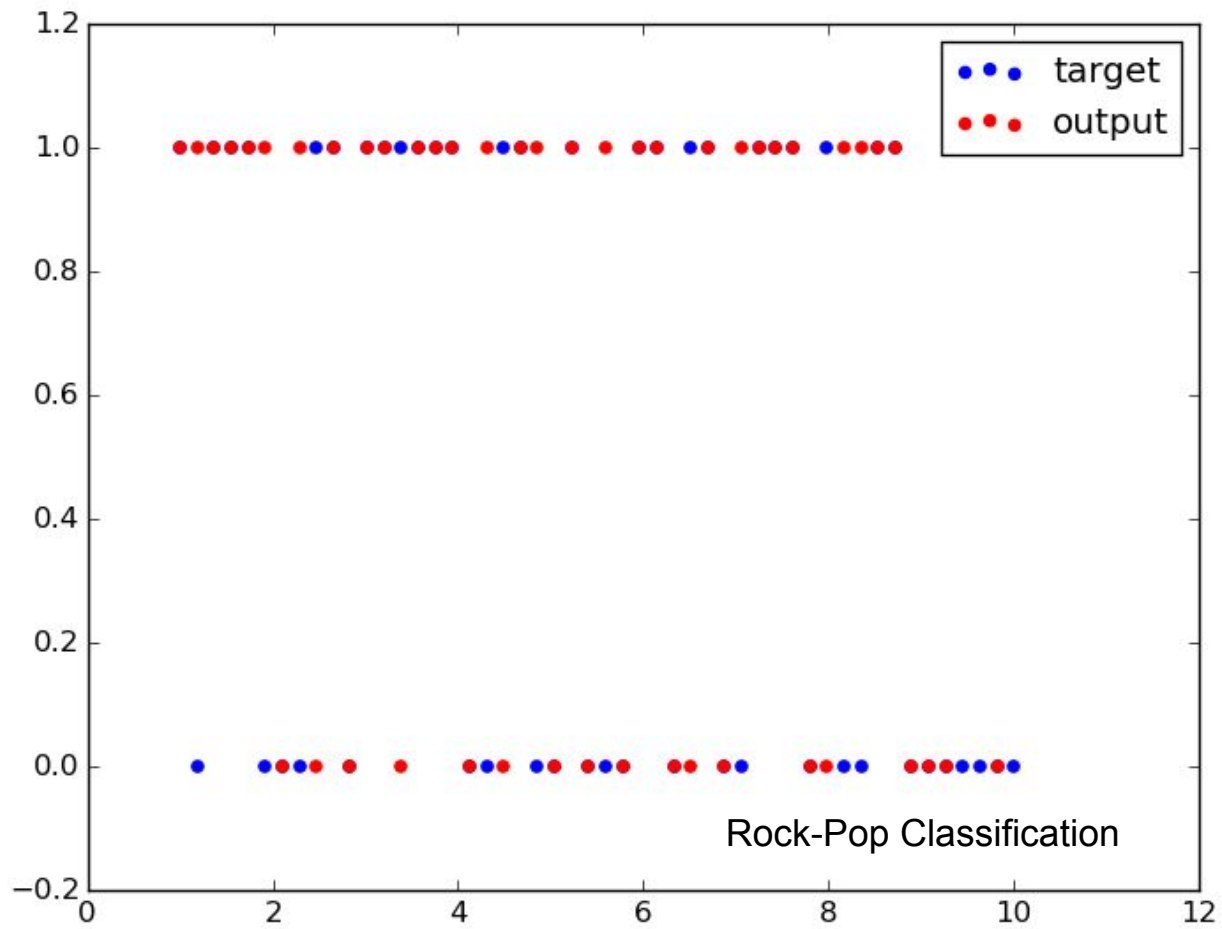
Outcomes of the project

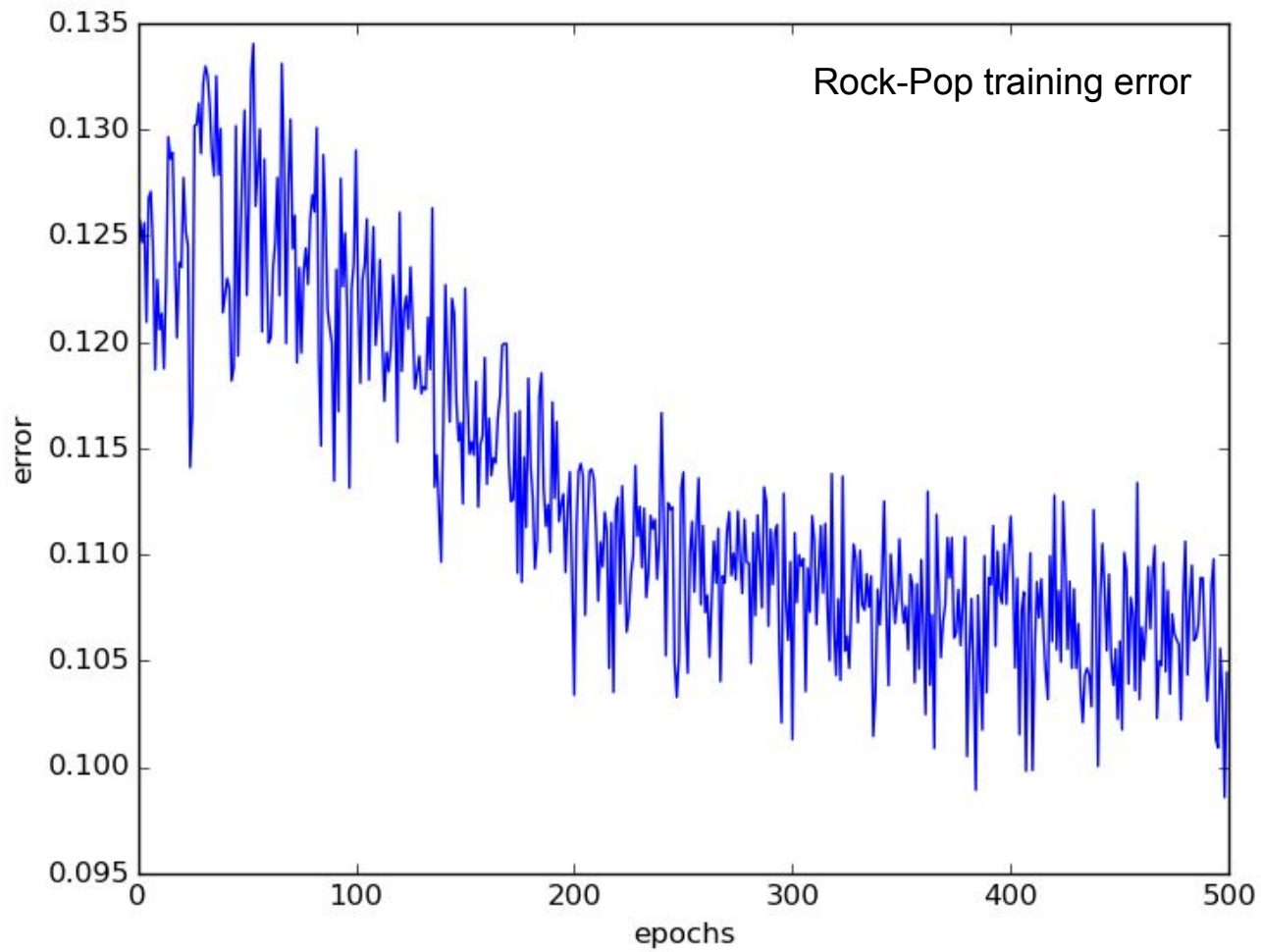
1. We use our program on different kinds of input files for classifying the music files into pop, classical, metal, rock etc. Currently we classify the files into 2, 3 and 4 genres with greater than 80% accuracy.
2. We find the total error after every iteration on training the Neural Net.
3. We calculate the sensitivity of our model.
4. We calculate specificity of our model.

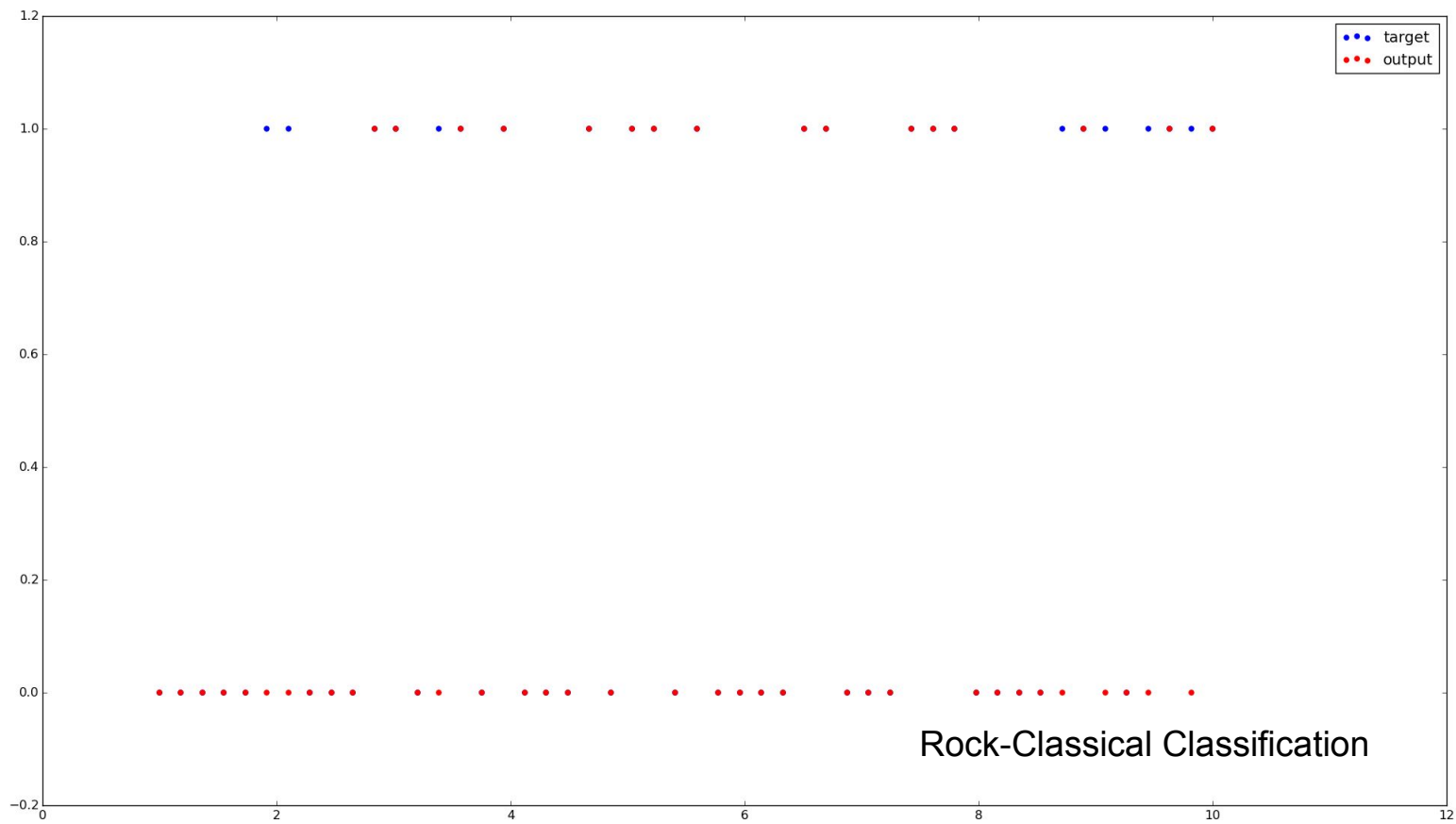
Screenshots

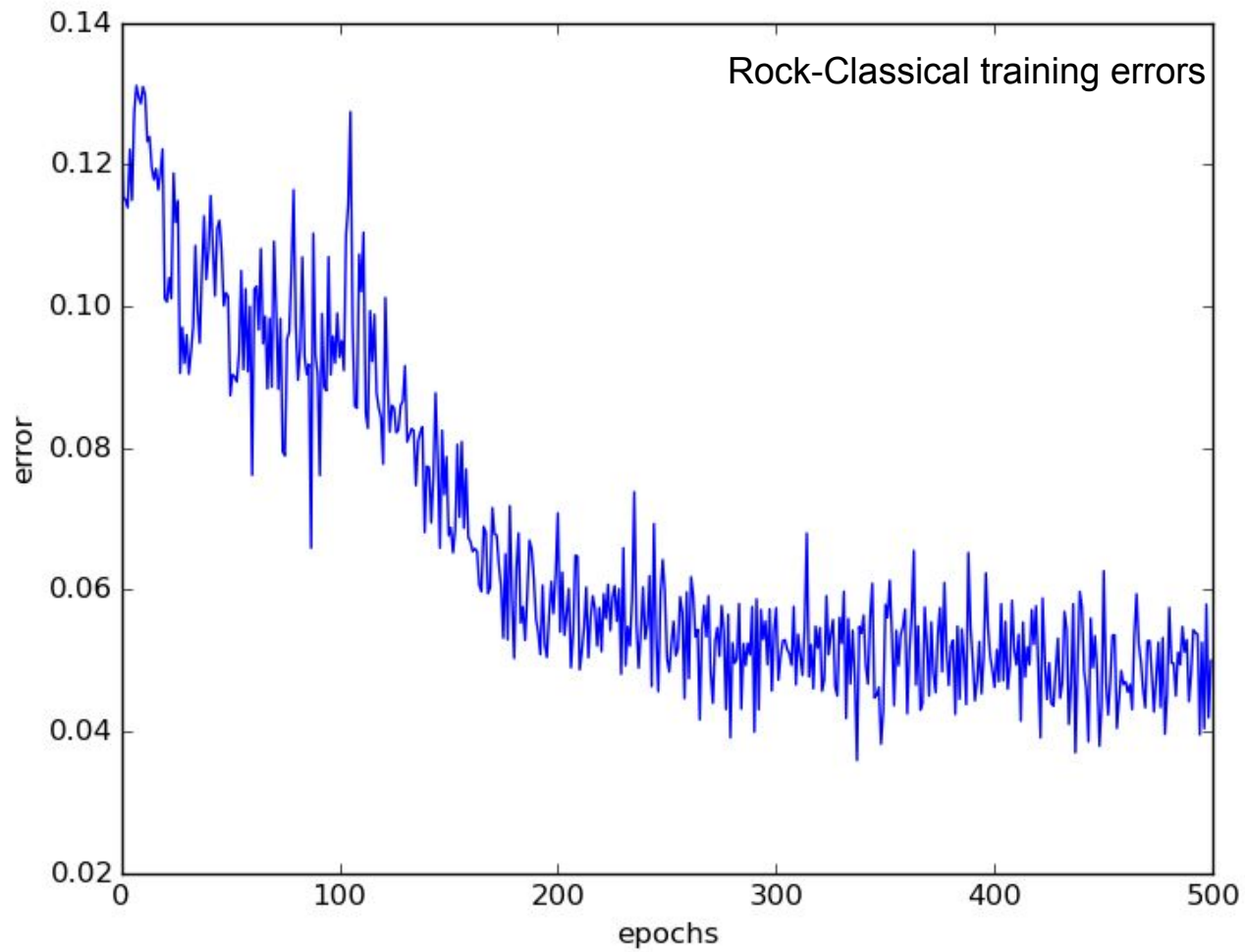
We have used 2 libraries: Pybrain and Neurolab.

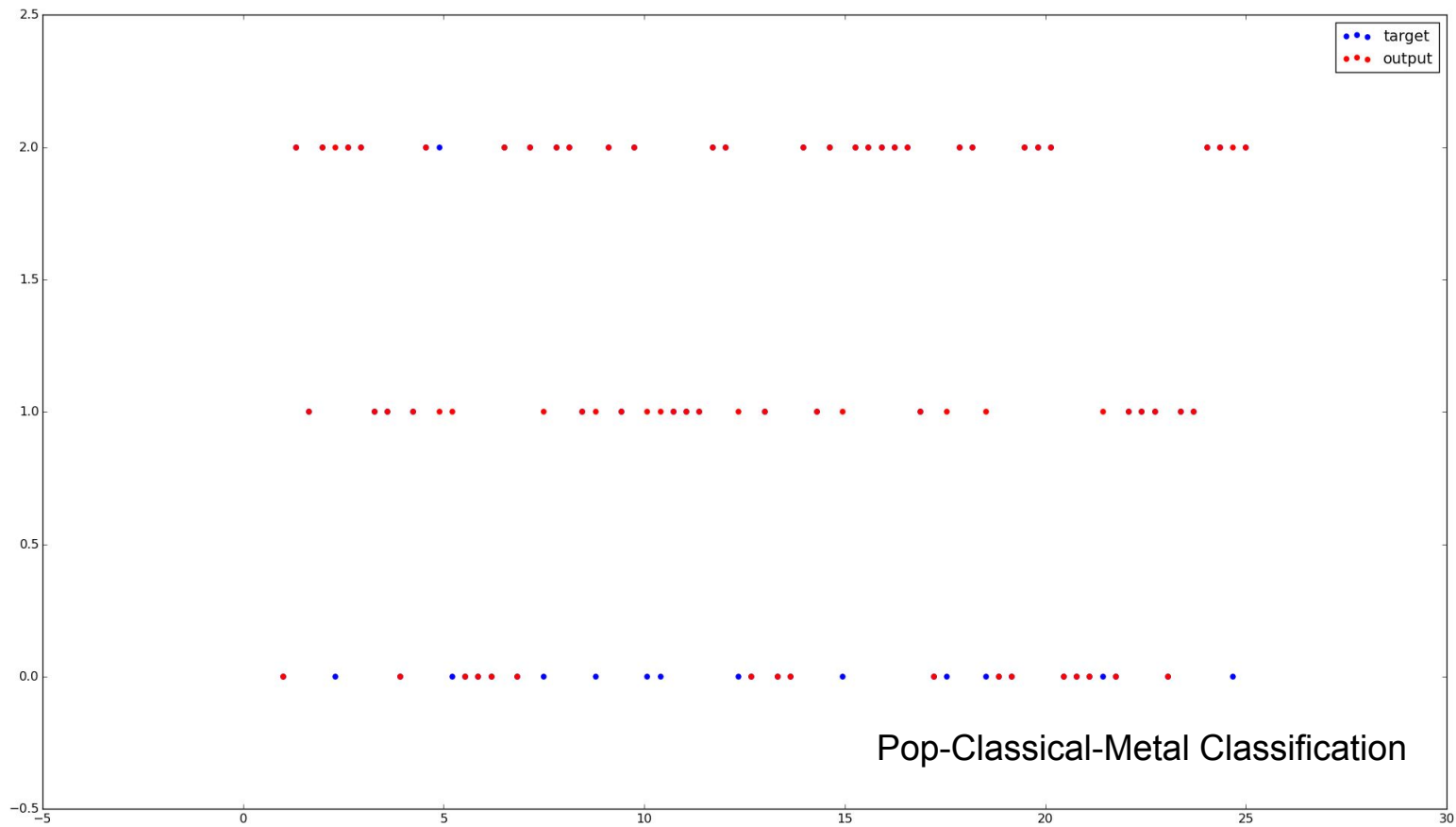
Some outputs and errors for Pybrain are:

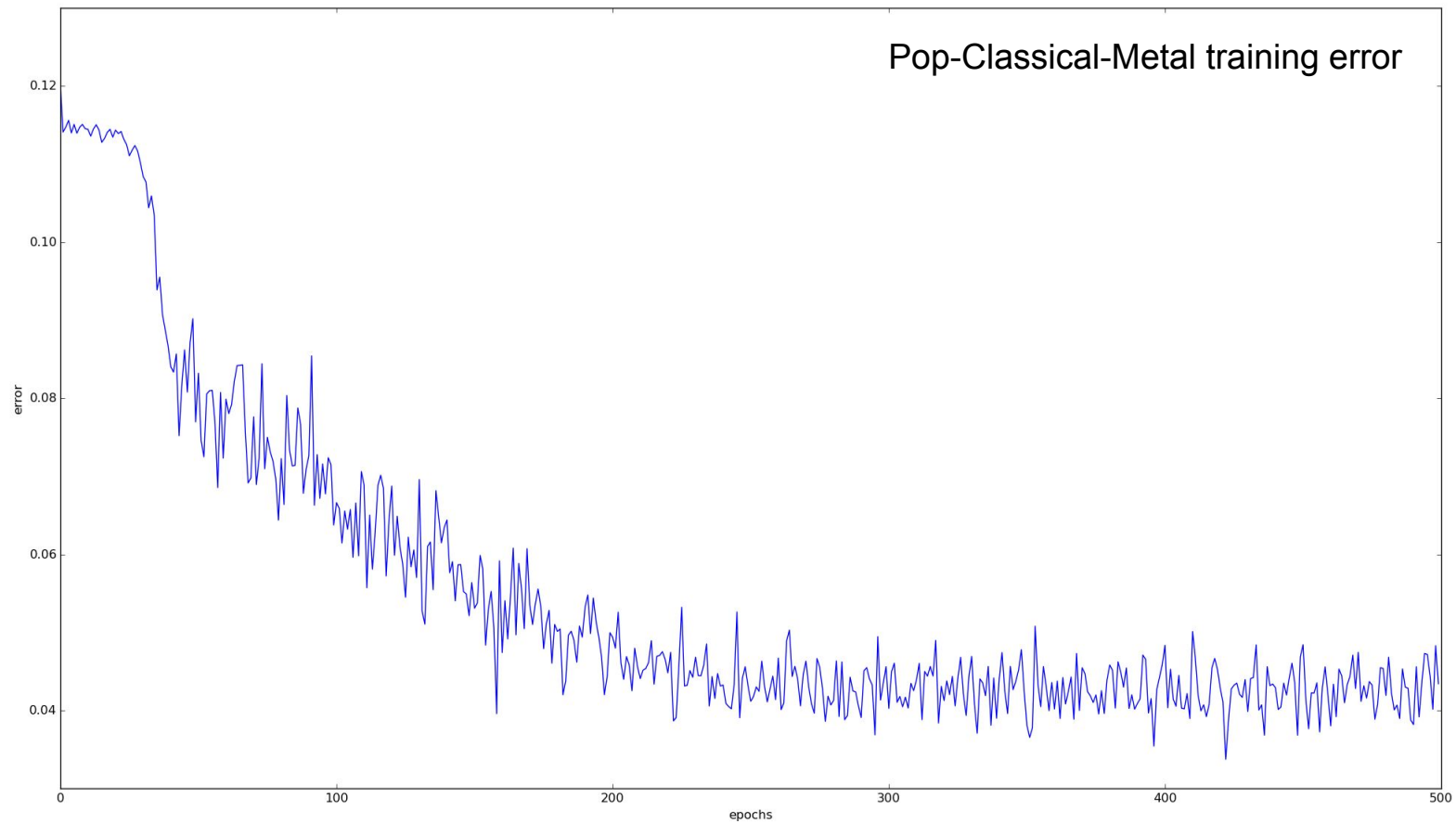


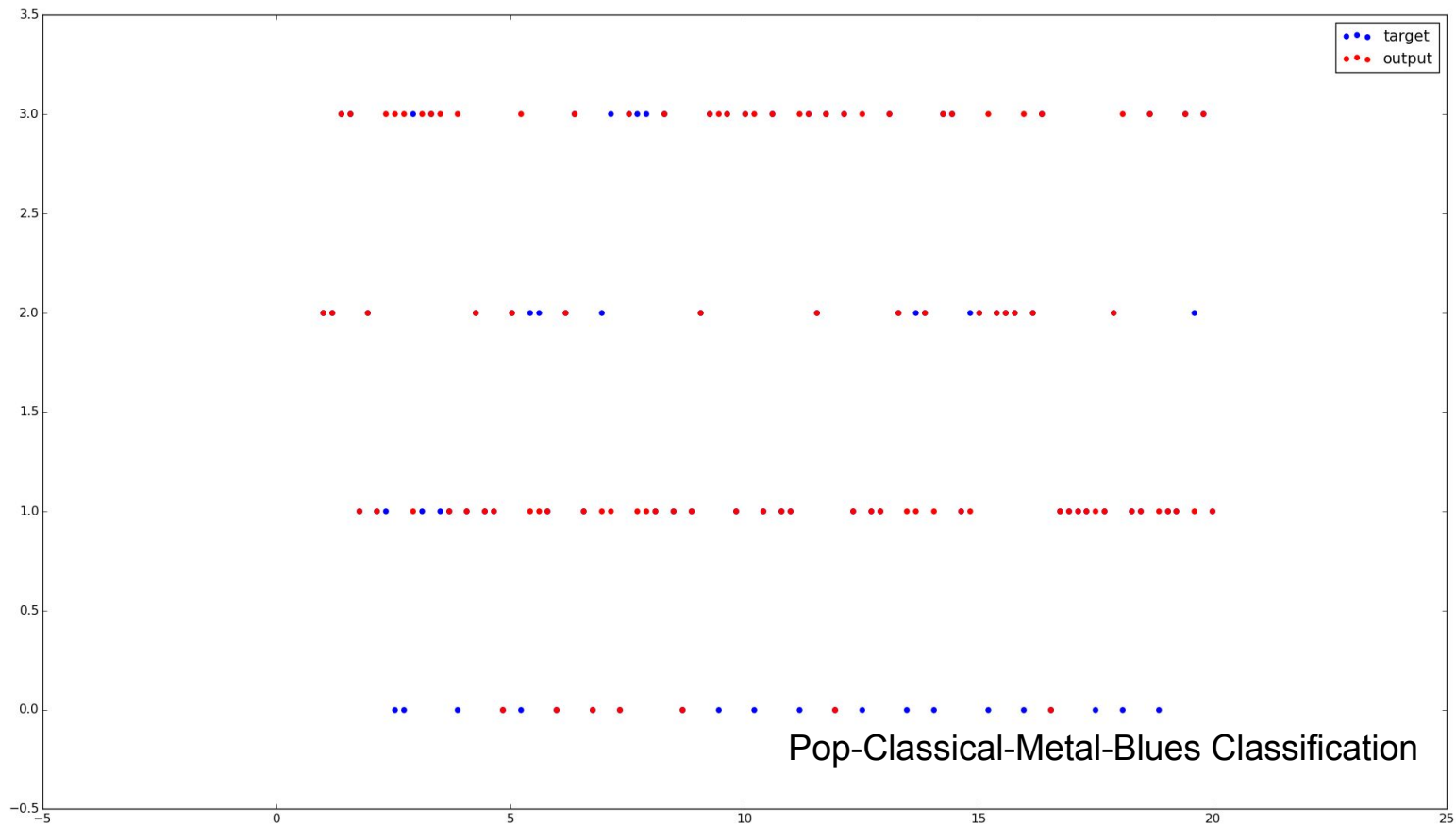


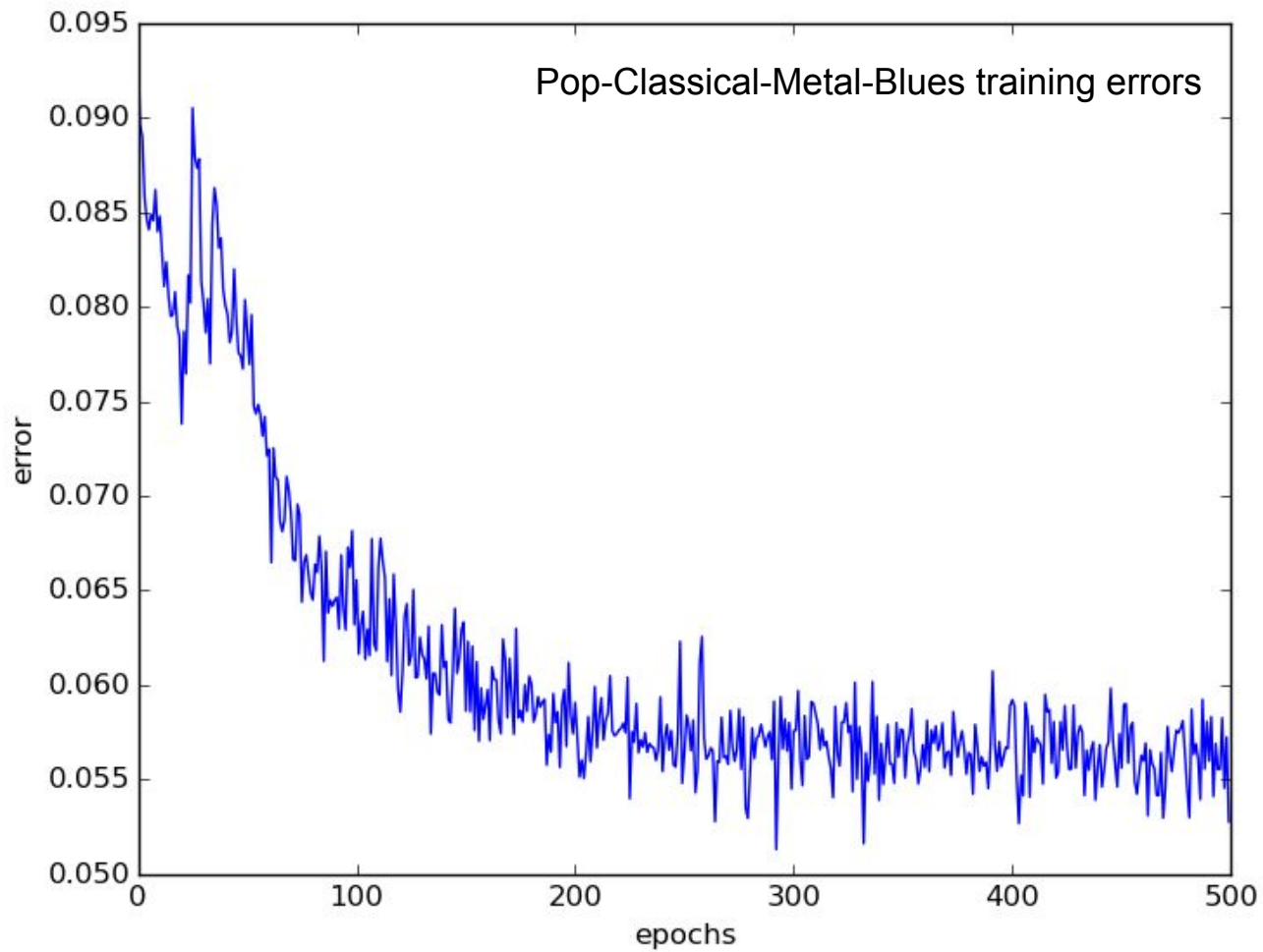






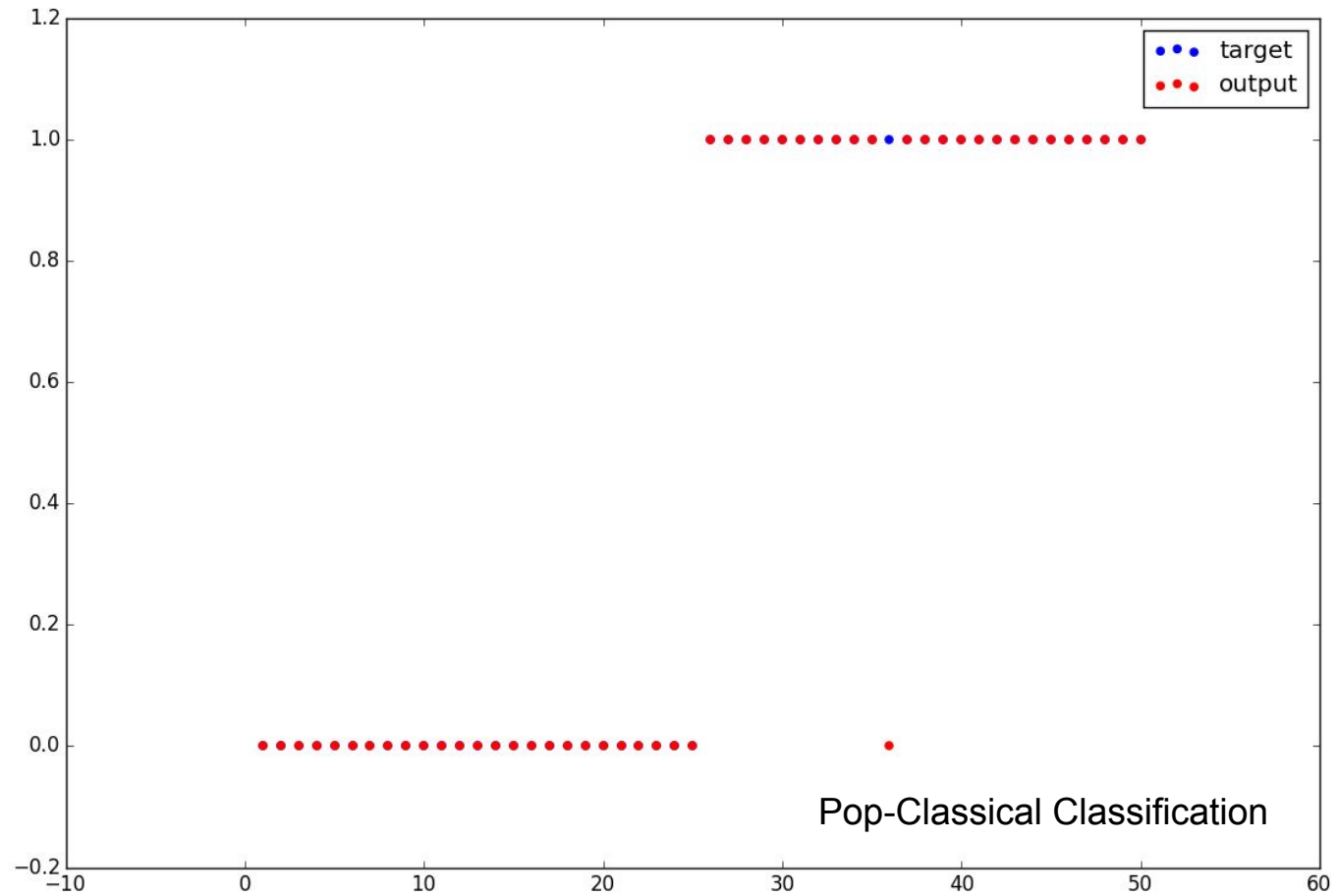


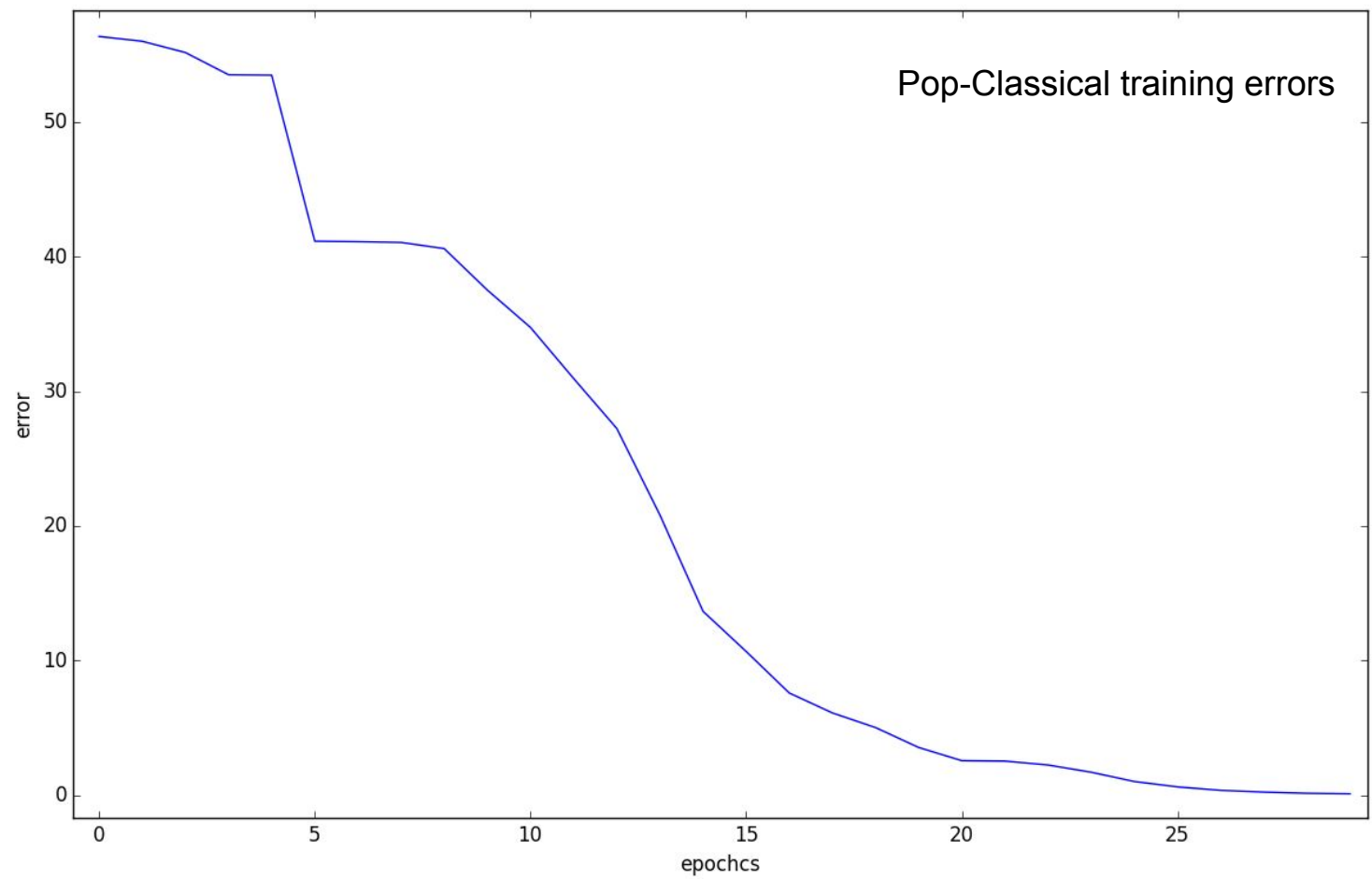


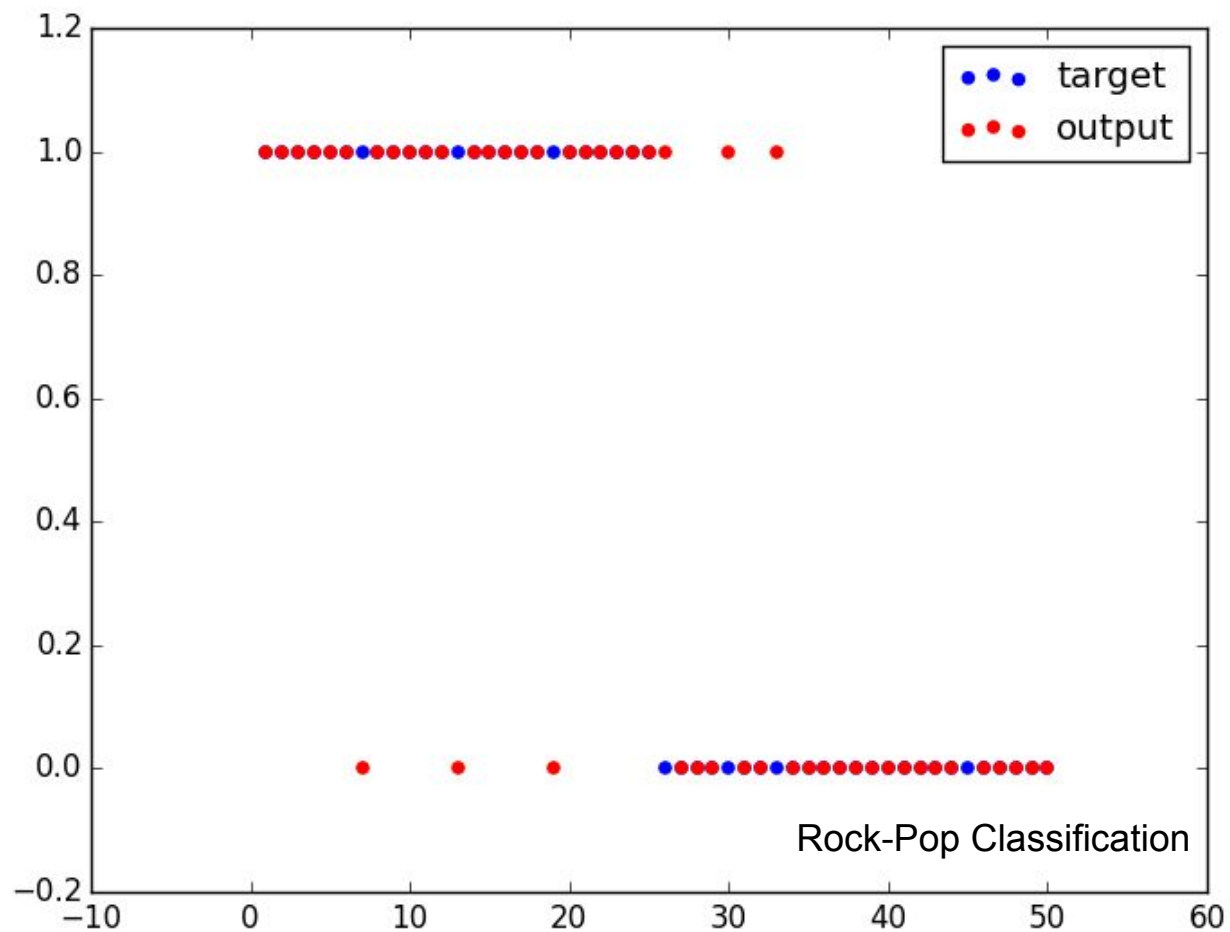


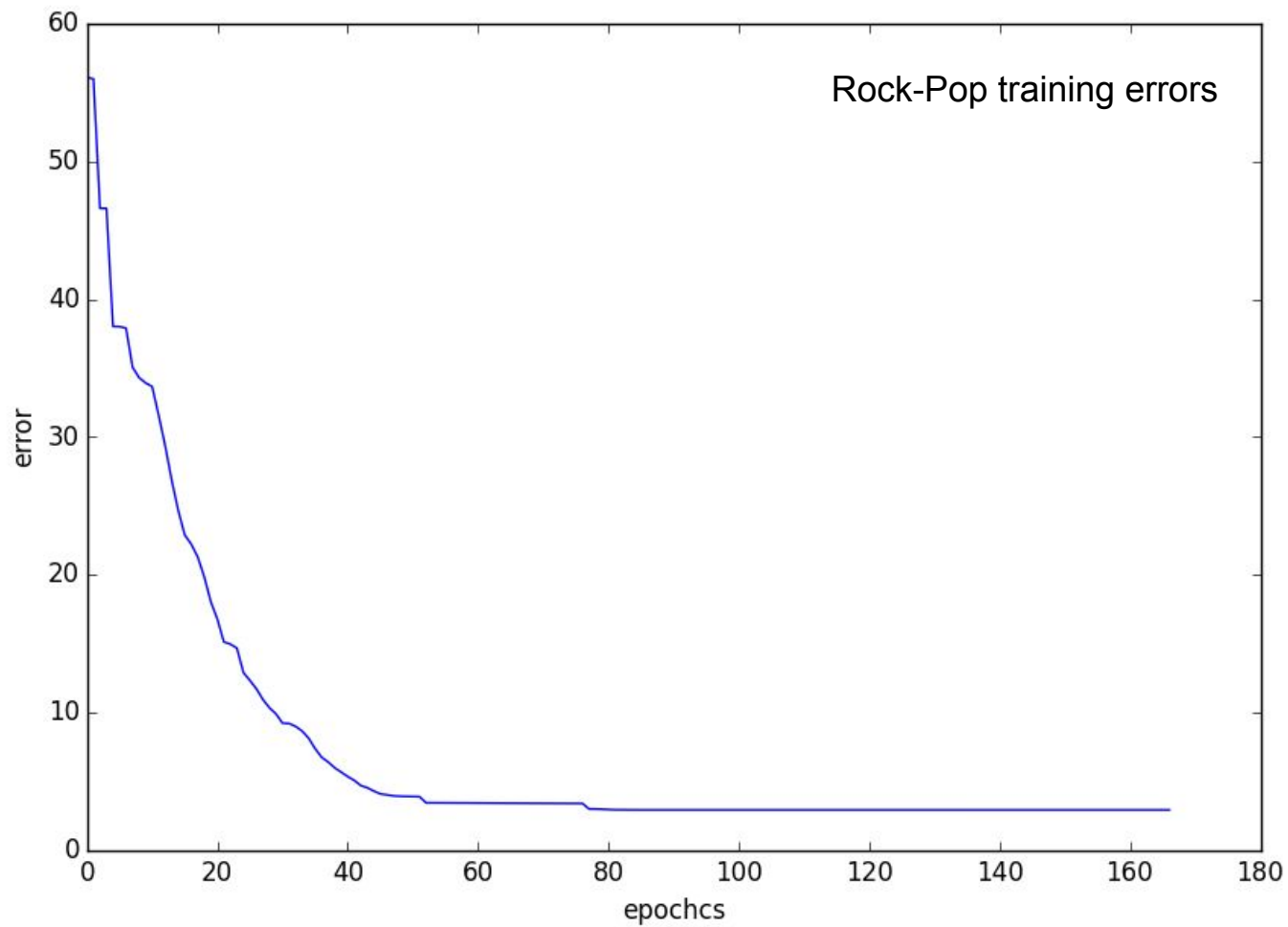
Screenshots

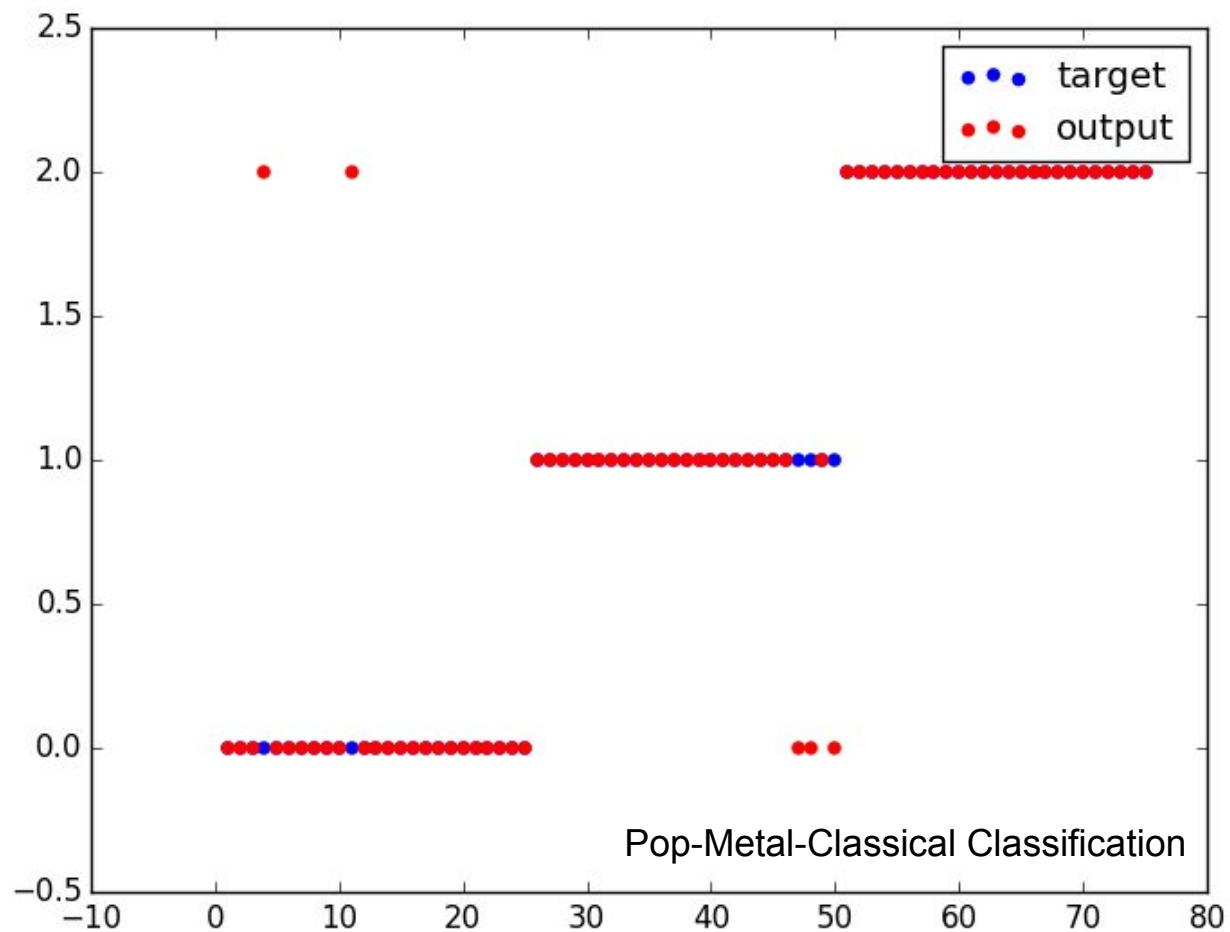
Some outputs and errors for Neurolab are:

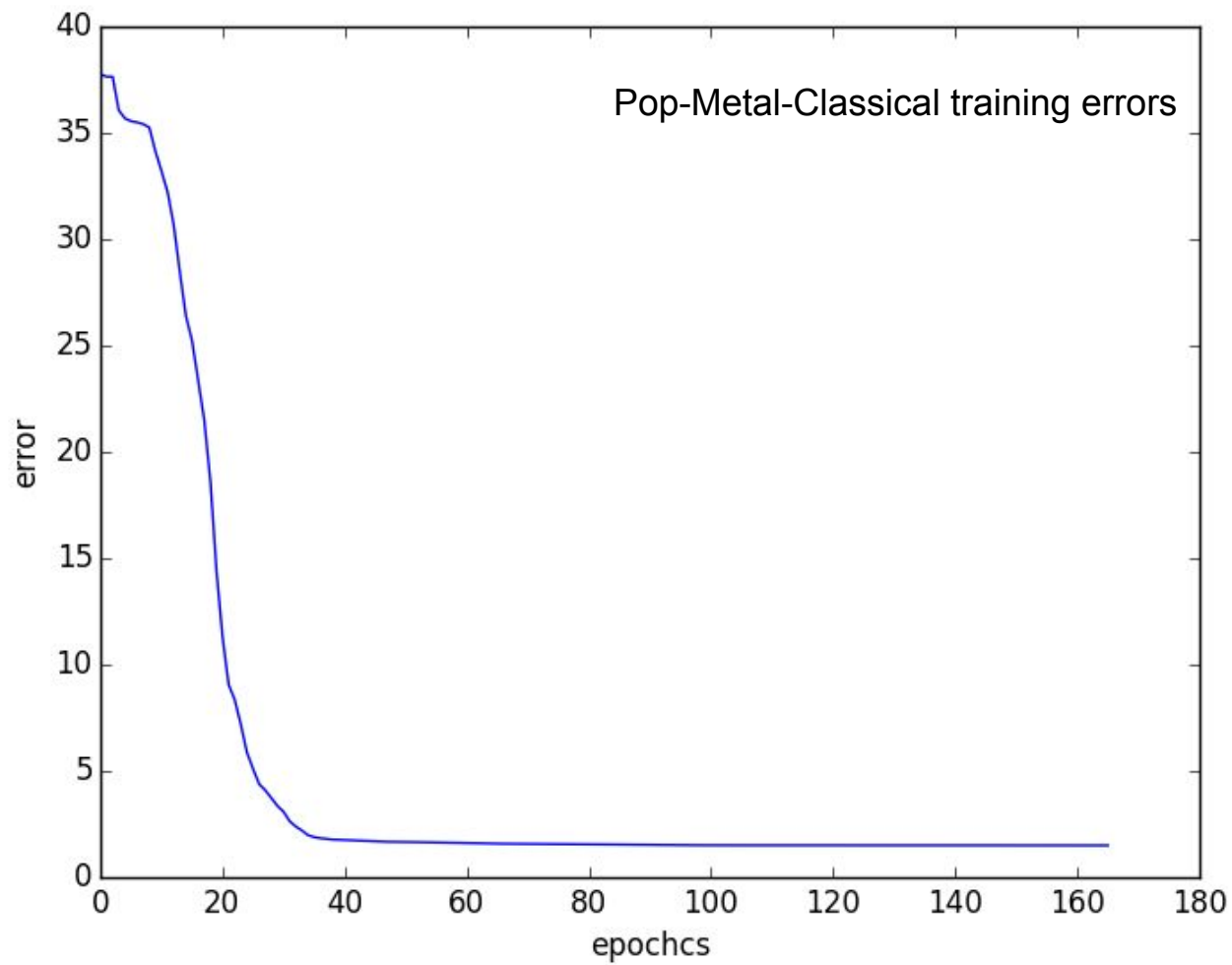


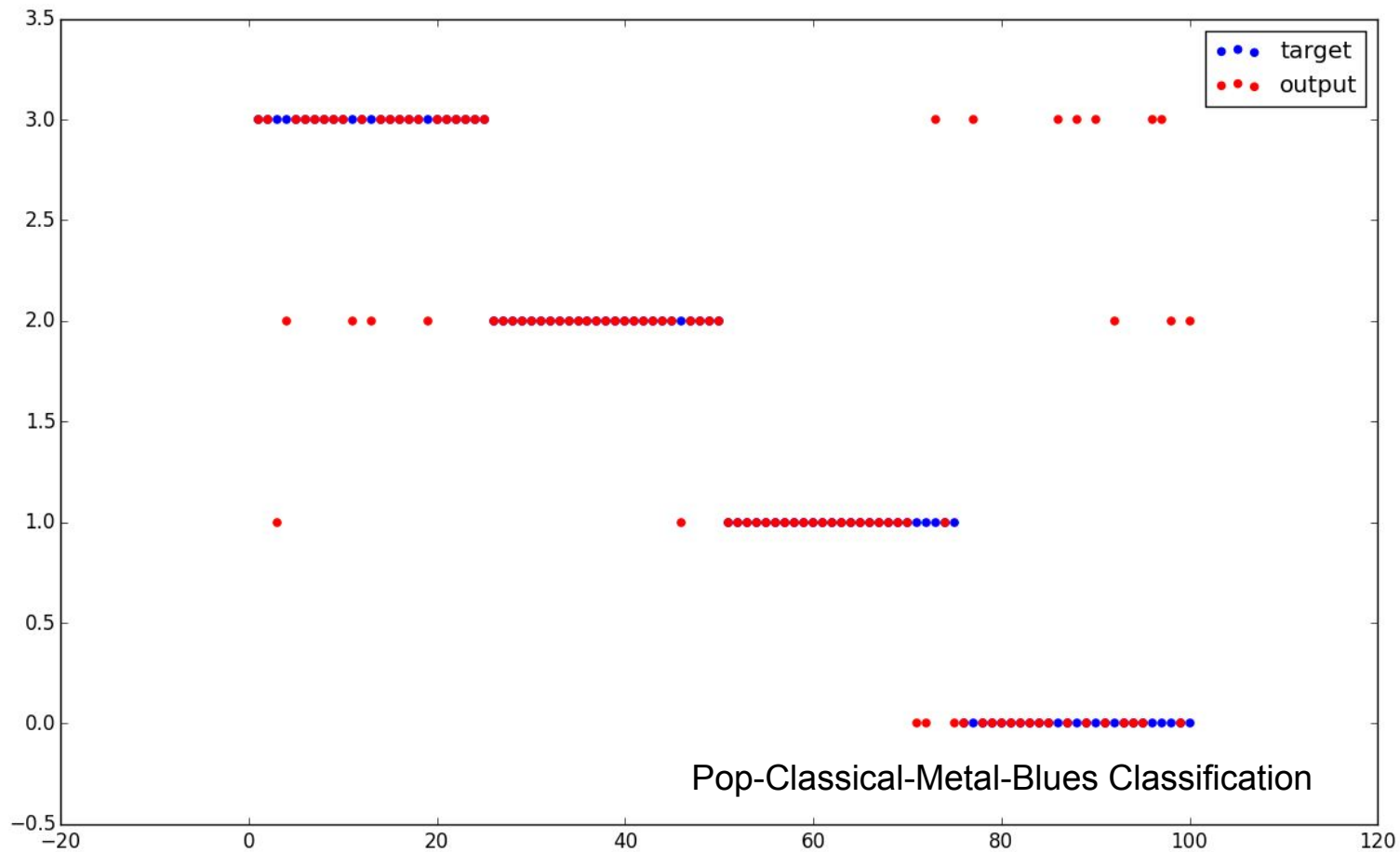


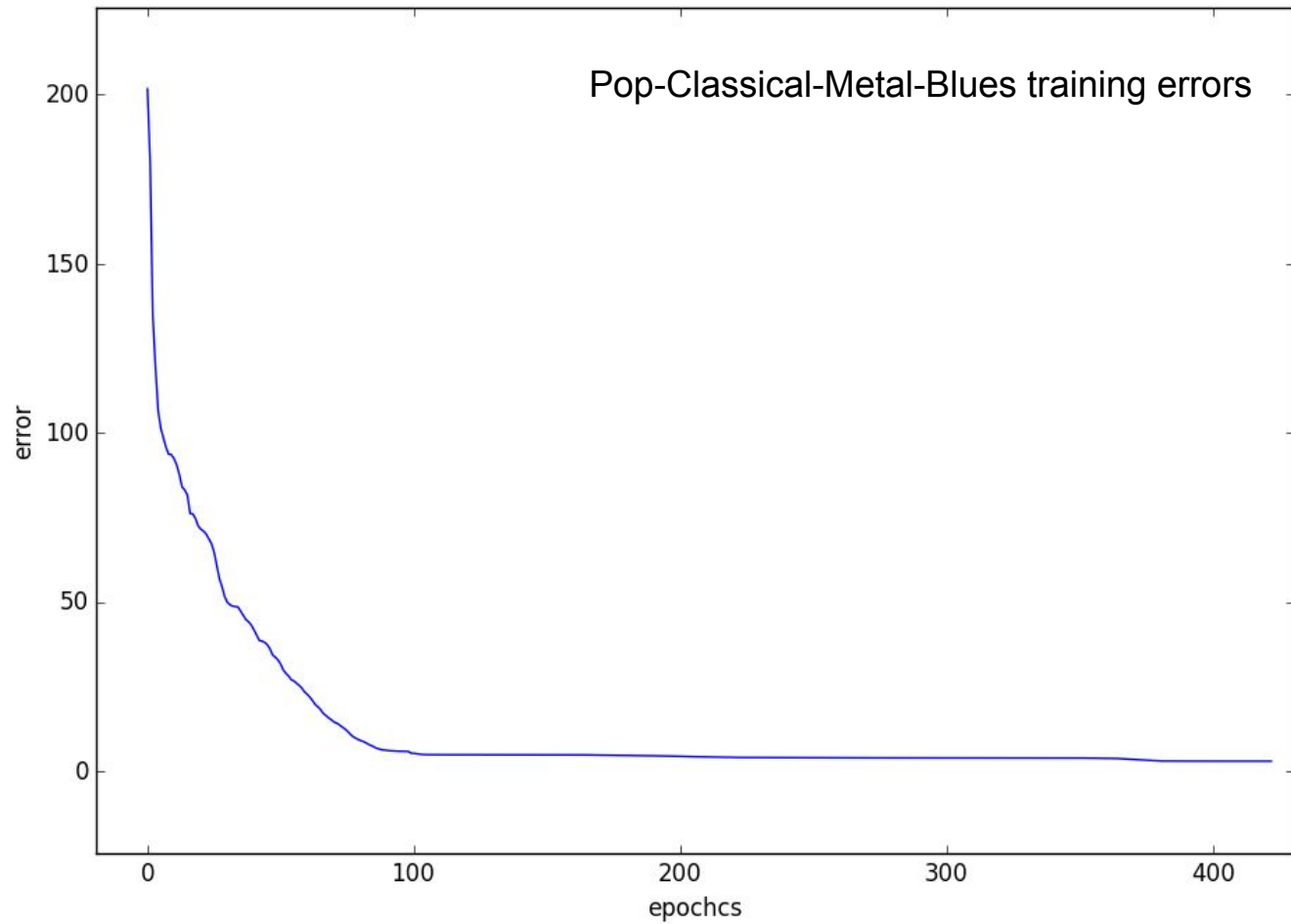












Code Overview

We have used music files, taken from GTZAN database, and extracted some features. These were used as inputs to our Neural Network. We have used 2 libraries to build a Neural Network, train and test the data. One is Pybrain and the other is Neurolab. By building a Neural Network and training it with the inputs it gives total error at each iteration. On testing it on the test data, it predicts the output.

Features Used : MFCC

We have used 52 features representing Mel-Frequency Cepstral Coefficients (MFCC) as input. MFC is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. MFCC is typically believed to encode timbral information, since it represents short-duration musical textures. Our input feature vector comprises of means and variance of several FFT frames.

Neural Network Model

For each genre we have a set of 100 sound clips with 52 input features. Training data and test data are split into 3:1 proportion. We have 1 hidden layer consisting of 20(pybrain) /12(neurolab) nodes and 1 output node for each genre. Our performance is measured against the following metrics:

1. Training error
2. Sensitivity for each genre : $tp/(tp+fn)$
3. Specificity for each genre : $tn/(fp+tn)$

Results for classification (Pybrain)

- 4 genre : Pop | Classical | Metal | Blues
 - specificity [0.83870967, 0.94736842,, 0.9285714, 0.93939393939393945]
 - sensitivity [1.0, 0.67441860465116277, 1.0, 0.58823529411764708]
- 3 genre: Pop | Metal | Classical
 - specificity [0.80327868852459017, 1.0, 0.96153846153846156]
 - sensitivity [1.0, 0.65789473684210531, 0.95652173913043481]
- 2 genre : Pop | Metal
 - specificity [0.93548387096774188, 0.94736842105263153]
 - sensitivity [0.94736842105263153, 0.93548387096774188]

Results for classification (Neurolab)

- 4 genre : Pop | Classical | Metal | Blues
 - sensitivity : [0.9733333333333334, 1.0, 0.9866666666666667, 0.84]
 - specificity: [0.56, 0.88, 1.0, 0.96]
- 3 genre: Pop | Metal | Classical
 - specificity [0.94, 1.0, 0.96]
 - sensitivity [0.92, 0.88, 1.0]
- 2 genre : Pop | Metal
 - specificity [0.96, 1.0]
 - sensitivity [1.0, 0.96]

Contribution

All of us have contributed equally. (33.33% each)

Future Suggestions

1. We have used only 30 second long sound clips for classification. Future work can involve longer length music which can have many variation of attributes in a single song.
2. In our current work, we have used only four broad genre, namely, rock, pop, classical and rock. These genres are clearly separated from each other. Future work may involve classification involving
 - a. More number of genres
 - b. Genres which have subtle differences (eg. country soul and country blues)

Learning from the project

1. The backpropagation algorithm performance is hindered by local minimas. Sometimes bad random initialization leads to local minimas out of which there is no coming out, no matter for how many epochs we train.
2. Learning rate is important for the performance of backpropagation trainer.
3. Choosing correct features also affects performance. Too few features lead to incorrect classification. Too many features introduce redundancy and make training more difficult.
4. Size of the network is also crucial. A very large network may lead to overfitting whereas a very small network may not have enough computational power to compute complex functions.