Read Me - Earthquake prediction
ECE 6254

May 10, 2019

The Earthquake - Time of Failure Dataset is 10GB and can be downloaded from the following link:

https://www.kaggle.com/c/LANL-Earthquake-Prediction/data

It includes a training set file and a test file in .csv format, used in our project.

Our Project codes are encapsulated in the following Python Notebook files:

- **rnn_model.ipynb**: Takes the raw training input file train.csv, to generate sequential features for the RNN model and training plots, and uses the raw segment .csv files in test.zip files for test predictions and MAE. Moreover, Kaggle GPU is used for training along with Keras Libraries to construct the recurrent neural network. Output is a submission.csv file for Kaggle Test set predictions, along with plots on the training set.

- **Feature-creation.ipynb**: Takes the raw training input file train.csv, to generate 156 features for each samples in the training data set and test data set. The output give us three files: test_features.csv which contains all the features for test data set, train_features.csv which contains all the features for the train data set and y_train.csv the time until failure of the train data set.

- **Feature-Selection.ipynb**: Feature Selection : Take as input the features matrices and it uses selection methods in order to calculate the importance of each features. Return as output three csv files containing the importance of each features using different selection method.

- **Models_most_important_features**: Fitting catboost and SVR using the top 15 important features.