Prediction of Timeseries and Text Classification (NLP)

Evaluation

Program 1 - Prediction / forecasting

- The goal is to reproduce the results obtained in paper

  Deep Learning and Gradient Boosting for Urban Environmental Noise Monitoring in Smart Cities

- Data
  - Provided by DataParkmeter.zip file (3 CSV file; one for each Parkmeter of the paper)
  - Preprocess data as explained
- Use TimeseriesGenerator to generate training or testing data
- Evaluate the prediction performance of the different deep learning models described in the paper to predict future noise levels (short / long term predict.)
- Once suitable predictions will be obtained try to
  - Use them in order to detect false data injection attacks as done in the paper
- You can also try to deal with pollution data particles...

Program 2 - Text classification

- The goal is to design a model able to classify text in two classes, predicting the sentiment polarity of Yelp Reviews
- Data are available in the NLP section of fast.ai datasets website
  - Dataset is rather large, you may use a subset of it
  - Labels are 1 or 2, you may change them
- You can implement a model described in the papers or your own one
  - Models like the ones seen for IMDB Sentiment classification
  - Transformer models; Transfer Learning
- Objectives
  - First design a model that you can train, without targeting the highest classification accuracy
  - Second tune the model to obtain the best possible classification performance