CS4361/5361 Machine Learning

Fall 2019

Lab 6 - Learning to predict sequences

Due Monday, November 25, 2019. Submit a paper copy by 5:00 p.m.

Email report to *olacfuentes@gmail.com*, include UTEP-ML2019 in the subject line.

The goal of sequence prediction is to predict future values of the variable of interest given past values of that variable. More precisely, at time t, we want to predict the values of $\langle x_{t+d}, x_{t+d+1}, ..., x_{t+d+p-1} \rangle$ given $\langle x_{t-b}, x_{t-b+1}, ..., x_{t-1} \rangle$, where d, b and p are positive constants, d is the delay (indicating how far into the future our predictions start) ,b is the lookback (indicating how far into the past an instance starts), and p is the length of the prediction window.

Your task is to write programs for sequence prediction. Specifically, you will use the solar dataset *xrp.npy*, which contains solar x-ray (first two columns) and proton fluxes (last six columns) measured every five minutes for a period of several years.

For our experiments, we will use a lookback of 144, corresponding to 12 hours, a delay of 36, corresponding to 3 hours, and a prediction window of 12, corresponding to 1 hour. Instead of predicting all the values in the prediction window, we will predict the mean of these values. Thus the first row in our data array would be *x[:144]* and the first target value would be *np.mean(x[180:192])*; the second row in our data array would be *x[1:145]* and the first target value would be *np.mean(x[181:193])*, and so on.

Implement and compare the following prediction algorithms. Use the first 1,300,000 examples as training and the rest as testing.

- A baseline, consisting of predicting the mean value of a row as the target value (that is, *predict(x[i]) = np.mean(x[i])*.
- LSTM using Keras
- CONV1D in Keras
- Any regressor from sklearn

As usual, write a report describing your work. For each of the 8 columns in the dataset, compare the performance of the algorithms with the baseline and comment about their relative performances.