Long short-term memory (LSTM) is an artificial recurrent neural network architecture. LSTM networks are well suited for classifying, processing and making predictions based on time series data.

In this project, you are supposed to classify the given dataset into 10 classes (digits 0 to 9) by the means of LSTM networks. The dataset contains time-series of Mel-frequency cepstrum coefficients (MFCCs), corresponding to spoken Arabic digits.

Number of instances (blocks): 8800
Number of attributes: 13

Each line in *Train_Arabic_Digit.txt* or *Test_Arabic_Digit.txt* represents 13 coefficients, separated by spaces.

Lines are organized into blocks, which are a set of 4-93 lines, separated by blank lines, and correspond to a single speech utterance of a spoken Arabic digit. Each spoken digit is a set of consecutive blocks.

In *Train_Arabic_Digit.txt*, there are 660 blocks for each spoken digit. Blocks 1-660 represent the spoken digit '0', blocks 661-1320 represent the spoken digit '1', and so on up to digit '9'.

In *Test_Arabic_Digit.txt*, digits '0' to '9' have 220 blocks for each one. Therefore, blocks 1-220 represent digit '0' , blocks 221-440 represent digit '1', and so on. Speakers in the test dataset are different from those in the train dataset.

Your tasks:

1. We want to use batch learning. Therefore, you should sort the training data by sequence length, and choose a mini-batch size such that sequences in a mini-batch have a similar length. Then you should pad the sequences in each mini-batch so that they have the same length. (Training process in some languages automatically add paddings to batches). This is necessary to prevent too much padding that may have negative effects.

2. After the data is preprocessed, build the LSTM network for classification and train it.

3. Repeat step 1 for the test data and then classify them.

4. Calculate the classification accuracy of the predictions.

5. Plot the loss and accuracy for test and train data over 500 iterations in separate figures.

Notes:

- Feel free to use any predefined functions.
- Due Date: Tir 26, 1399.
- Pay extra attention to the due date.
- Be advised that all submissions after the deadline **would not be graded**.
- Prepare a complete report.
- Email your files as a folder in this format (HW#_student#_name_family.zip).
  Email: hashemi.mahshadd@gmail.com

Kind Regards