Assignments due on 28/03/2022

Assignment 7a

a) In this assignment, we will develop a language model that will suggest the rest of the sentence for given input.
b) The idea is to simulate Google's suggestions when we try to search.
c) For example: if you input “I eat”, output suggestion will be “rice in a restaurant”. The suggestion would be based on the training with your hundreds of sentences collected from Wiki.
d) Your maximum sentence length is 8 words. So in this model, you will use LSTM eight times to capture the recurrence. One hot encoding of a word would be input to the LSTM cells.
e) During the training, if the sentence is less than 8 words, you will use EOL to fill up the sentence. EOL will have a unique one hot encoding.
f) So training with this language model will help the LSTM learn the word sequences it has seen in wiki documents.
    a. At first, read all sentences from wiki, clip them in to 8 word sentences. Add EOL if necessary. Find unique word. Then build your vocabulary and calculate one hot encoding for each word.
    b. Build the language model with LSTM cell for eight recurrences.
    c. Use the sentences for training.
    d. Example of training sentences would be supplied to the language model as follows:
        i. I drink tea eol eol eol eol eol
        ii. Bangladesh has been playing cricket since 2000 eol
        iii. I drink coffee in café eol eol eol
g) After training, take a test data (partial sentences) and test the prediction accuracy with the ground truth.

Assignment 7b

a) Write a small program that ask you to input a word or part of the sentence.
b) The program will suggest the rest of the sentence. Then the program will ask for next input.
c) Example:
    1. Input: I play
    2. Output: cricket everyday eol eol eol eol
    3. Input: I live
    4. Output: in Bangladesh eol eol eol eol