NLP Assignment-III   Date: 22-5-2020 Time: 4 -5:30   PM

Instructions to students:

Assignment Question:

Consider the following toy example training data:
<s> I am Sam </s>
<s> Sam I am </s>
<s> Sam I like </s>
<s> Sam I do like </s>
<s> do I like Sam </s>
Assume that we use a bigram language model based on the above training data.

1. What is the most probable next word predicted by the model for the following word sequences? Show mathematically.
   (a) *<s> Sam . . .*   (b) *<s> Sam I do . . .*
   (c) <s> *Sam I am . . .*   (d) *<s> do I . . .*
   (e) <s> *Sam like . . .*

                                                                                              (2.5)

2. Calculate the probabilities of the following sequences after applying Laplacian Smoothing
   (a) *<s> do Sam I like*   (b) <s> *Sam do I like*
   Which of the two sequences is more probable according to the Language Model (LM)?

                                                                                              (2.5)