

Instructions to students:

Students must write the answers in their own handwriting on a white sheet of paper. The details to be mandatorily written on the answer sheet are: **Name, Semester, Section, Roll Number, Registration Number, Course name, Signature with date.**

On completion of answering the assignment, students need to scan/image (using device of your choice) all the answer sheet/s in sequence and **save it with file name of their registration number in PDF/ Image format.**

Upload the above pdf/image file containing the answers to MS Teams Assignment platform within the given schedule.

CLICK ON "ADD WORK" OPTION, ATTACH THE PDF/IMAGE AND THEN CLICK ON "TURN IN".

Student to contact faculty member concerned through mail/phone in case of any difficulty faced by them during the assessment process.

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)