Date: 09-09-2024   **Quiz -1**   **Time allowed: 25 mins**

NLP-B   **Total Marks: 18**

Q1. Identify the root word and inflectional/derivational affixes in the following words. (mention each root word, and its inflectional or derivational affix, mention the one that is applicable)   (9)

1. Boxes (root:_____ ; inflectional/derivational: _____)

2. terrorize

3. Unaccountability

4. Unjustifiably

5. Spoken

6. Misfortune

7. بے پناہ

8. لا حاصل

9. لاپروابی

Q2. Suppose you have a language model trained on a corpus of English text using unigram probabilities. How would you expect the perplexity of this model to differ between two sentences: "the cat chased the mouse" and "the mouse chased the cat"?   (1)

    a. The perplexity on "the cat chased the mouse" is greater.
    b. The perplexity on "the mouse chased the cat" is greater.
    c. The perplexities are equal.

Q3. Can stemming or lemmatization help reduce the impact of OOV words? Explain by giving 3 examples.   (3)

Q4. Assume you have the following training corpus, include </s> in your corpus like any other token. Perform all steps.   (5)

<s> i am from Vellore </s>
<s> i am a teacher </s>
<s> students are good and are from various cities</s>
<s> students from Vellore do engineering and medical</s>

Find the unigram and Bigram probability of the given test sentence, including </s> as a token. Calculate the perplexity of these both as well.

Test data:
<s> students are from Vellore </s>