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

NLP-A   **Total Marks: 18**

Q1. Identify the different types of morphemes (bound & free, or reduplication etc.) in the following words:

(6)

1) Pharmaceuticals

2) outrageous

3) sunscreen

4) کتابدار

5) محافظت

6) postpone

Q2. Mention a few techniques to handle OOV words. Explain with reference to few examples.   (5)

Q3. How do BPE and maxmatch differ? and which algorithm is a better tokenizer than the other? Explain with an example.   (2)

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

<s> I love eating thai and asian </s>
<s> My favorite food is sushi and ramen </s>
<s> Eating out is fun and exciting </s>
<s> Fast food is my go to meal </s>
Find the ***Unigram and Bigram*** probability of the given test sentence.
Calculate the perplexity of these both as well.

Test Sentence:
<s> I love eating ramen </s>