CSE556: NLP

Mid-Sem

Duration: 1 hour

Please submit before the deadline. There will be two marks deduction for late submission up to 11:05 am. After that your response will not be evaluated.

Total marks: 40

Note:

- It's a pen-and-paper-based exam. **On every sheet, please write your name in BOLD LETTERS.** We will not evaluate if you fail to do so.
- There will not be any extension and also no exemption will be given to anyone.
- Each student has to switch on his/her camera during the exam.

1. Given a tag set T = [NN (noun), VB (verb), JJ (adjective), RB (adverb), IN (preposition), PRP (pronoun), O (Others)], find the correct PoS tag sequence.
   a. I shot an elephant in my pajamas. How it got into my pajamas I will never know.
   [8 marks, no partial marking]
   b. The complex houses married and single soldiers and their families.
   [4 marks, no partial marking]

   Note:
   - Response should be in the format <word1_tag1> <word2_tag2> …. <wordN_tagN>
   - A single mistake will incur zero mark.

2. Given the following probabilities, compute the state sequence for the observation "RRR" using the Viterbii algorithm. **Show the decoding tree**. **[20]**

Transition probabilities

	B1	B2	B3
B1	0.1	0.4	0.5
B2	0.6	0.2	0.2
B3	0.3	0.4	0.3

Emission probabilities

	R	G	B
B1	0.3	0.5	0.2
B2	0.1	0.4	0.5
B3	0.6	0.1	0.3

Note: Initial transition probabilities is given as 0.33 each

3. Design FST to convert intermediate forms of the words (Sheep and Fish) to their lexical forms. **[8]**