

# Homework 2

CS312 - Spring 2024

This homework is to help students practice the part-of-speech tagging and the grammar parsing.

**Question 1:** Suppose we have a dataset of texts and annotations as the following

their(N) raises(N)  
he(N) raises(V)  
my(N) purses(N)

- Find all parameter values (i.e. transition probabilities and emission probabilities) of a HMM built from this dataset. Note that you need to add the starting state  $\langle s \rangle$  into the transition probabilities.
- Find an example including a sentence and its tags (both words and tags are from the dataset above) that the HMM in (a) assigns a zero probability.
- Use the Viterbi algorithm to find the tags for the sentence “he raises purses”, you need to show the table where you calculate the  $v_t(j)$  as in the lecture/textbook.

**Question 2:** Suppose we have a PCFG in which the nonterminal set is  $\{NP, PP, NNS, CC, P\}$ , the terminal set is  $\{cats, and, in\}$ , the start symbol is NP, and the rules are

$NP \rightarrow NP CC NP [0.3]$   
 $NP \rightarrow NP PP [0.3]$   
 $NP \rightarrow NNS [0.4]$   
 $PP \rightarrow P NP [1.0]$   
 $NNS \rightarrow cats [1.0]$   
 $CC \rightarrow and [1.0]$   
 $P \rightarrow in [1.0]$

How many valid syntactic parses (i.e. non-zero probability using the grammar above) of the sentence “cats and cats” and the sentence “cats and cats in cats”? You need to show how you find this number of parses.

**Question 3:** Install Stanza and use the constituency/dependency parsers of the Stanford CoreNLP

<https://stanfordnlp.github.io/stanza/constituency.html>

<https://stanfordnlp.github.io/stanza/depparse.html>

- Draw the parsing trees of the sentences “I ate spaghetti with chopsticks” and “I ate spaghetti with meatballs” using the constituency/dependency parser. Do you think the results are correct or make sense?

- b) Find two sentences that you think that they may be ambiguous and then show their constituency/dependency parsing trees using Stanza. Explain if the results are the ones you expected.
- c) Find a sentence that has at least five children for a single word in its dependency parse. You may try different sentences first to have some observations and ideas. Then you can search for the sentence.