CSE 325/425 (Spring 2021) Homework 5

Due on 11:55pm, Apr 21, 2021

Grading: All questions have the same points (25 each). We will randomly grade some of the questions.

Submitting: Only electronic submissions on Coursesite are accepted. You can handwrite your answers on papers and then scan them to images. If you need to plot figures using a computer, the plotted files should be saved and included in the submitted pdf file. Submit a single pdf file named

<Your LIN>HW5.pdf

Other format will not be accepted.

Questions:

1. Assume the following PCFG:

   ```
   S -> a (with probability 1/3)
   S -> Sa (with probability 2/3)
   ```

   where a is a terminal and S is the only non-terminal. Prove that the total probability of all sentences generated by the PCFG is 1.
   (*Hints: you will need to find all sentences, each of which has a probability of being generated. Then sum up the probabilities and show that the sum is 1.*)

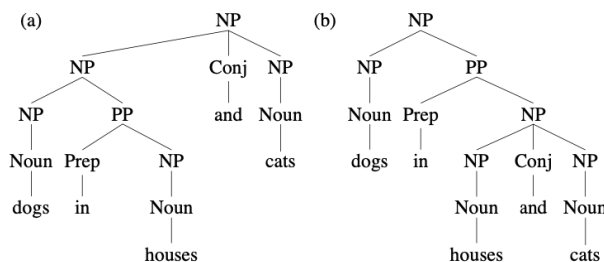2. Prove that the two parsing trees will have the same probability under the same PCFG.

Figure 1: Two parsing trees for the input *"dogs in houses and cats"*

3. Analyze the time complexity of the CYK algorithm running on a sentence of length m with a CFG with N non-terminals. Use the Big-O notation. Answers on specific sentence or CFG will not be accepted. (*Hints: run the CYK algorithm on some simple sentence and CFG to get a sense of how much computation you need.*)

4. Explain in what order should the CYK matrix be filled out when calculating the outside probabilities using dynamic programming.

5. In the algorithm that finds the most likely parse tree for a sentence, explain where and what information needs to be stored to reconstruct the tree.