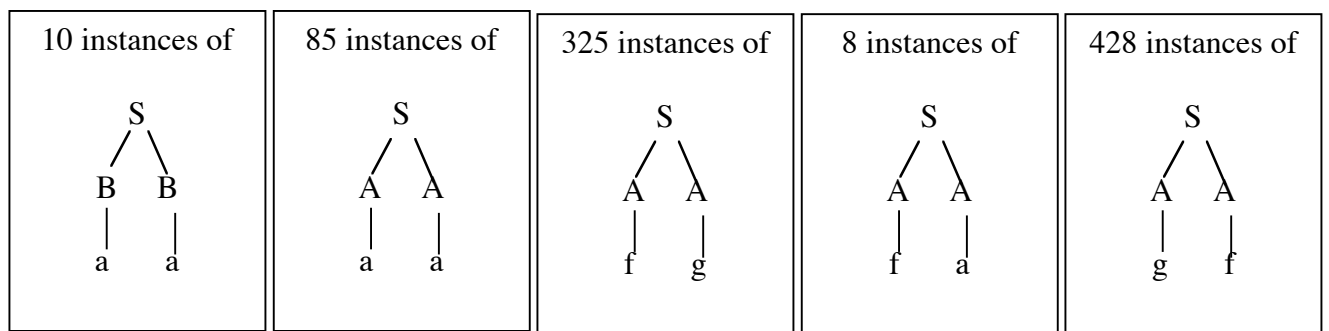


CSC 248/448 Makeup Assignment/ Review for Exam

Due: at the start of class, Dec 9. No late assignments accepted as we will discuss these questions in the class.

If you hand in this assignment, then the grade can be used to replace 75% of a previous assignment grade. For instance, if you missed an assignment, you can earn up to 75% on that missed assignment grade. Likewise, if you did poorly on an assignment, this can allow you to improve your grade on that assignment.

1. PCFG's tend to have a bias towards using nodes that occur in the right hand side of just a few rules. To see this, consider the following training corpus.



- What is the PCFG derived from this corpus using the MLE.
- What is the most likely parse for the input “a a”.
- Is this reasonable given the data? (Justify your answer either way)
- Would answer of the more powerful probabilistic models discussed in class produce a better answer. Discuss each and state how they might help, or why they don't.

2. Machine Translation

- In a translation task, we estimate $\text{argmax}_e P(e | f)$ by rewriting using Bayes rule to produce $\text{argmax}_e P(e) P(f | e)$. Why do we do this. What would happen if we simply directly modeled $P(e | f)$ and did without the language model. Give one of more reasons why the rewritten model should produce better English translations.
- Consider a simple translation model in which we simply map words from one language to another, and allow words in any position to be equally likely to appear in any position in the translation. In effect, the translation model makes every ordering of the translated words equally likely. For example, the French sentence *Je T'aime* would produce equally likely sequences *I you love*, *I love you*, *you I love*, *love I you*, and *love you I*. Could we still produce reasonable translations using this translation model. Why? In what cases would it tend to fail? Give an example.

3. Parsing

a) Extend the probabilistic CKY algorithm so that it records the tree as it build constituents. Give pseudo code in sufficient detail so that the program would be easy to write.

b) Hand-simulate your algorithm and produce the chart for the sentence *astronomers saw stars with ears* given the grammar:

S → NP VP	1.0	NP → NP NP	.1
PP → P NP	1.0	NP → NP PP	.3
VP → V NP	.7	NP → astronomers	.1
VP → VP PP	.3	NP → ears	.18
P → with	1.0	NP → saw	.04
V → saw	.9	NP → stars	.18
V → stars	.1	NP → telescopes	.1

c) Draw the highest probability parse tree and give its probability.

d) How would you modify the CKY algorithm if you wanted the parser to return the two most likely parses.

e) What is the second most likely parse of the sentence in part (b)?

4. Vector Space Models

Due to the calculations required in this problem, you probably want to use a spreadsheet, or write a program to answer the questions.

Assume we have the following information about certain words (A, B, C, D, E and F) occurring in 10 different documents.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
A	1		2	1			1			
B	1		1	3	1					
C		2					1			1
D	1	1		1		1	1			1
E	2	3		4	1	2		3	3	1
F	1	1	4	1	1	3	2		1	

Word counts (term frequencies) in documents (tf_i)

a) Describing and justifying all the steps, determine a measure that tells us which document is most similar to D2. What is your answer given your method.

b) Say we need to represent the documents using only a 5 dimensional vector. Describe carefully a method for determining what these five values represent and compute a good vector representation for this set of documents.

c) Compare the five- and six-element vector representations of the document by checking which document is closest to a “query” consisting of A C E for each.