Midterm, COMS 4705

Name:

4	8	10	10	15	15

Good luck!

Part #1 4 points

Say we have a language model with vocabulary $\mathcal{V} = \{\text{the, a, dog, cat}\}$, and $p(x_1 \ldots x_n) = 0.5^n \times g(x_1 \ldots x_n, n)$ for some function $g(x_1 \ldots x_n, n)$, where $x_i \in \mathcal{V}$ for $i = 1 \ldots (n-1)$, and $x_n = \text{STOP}$. Which of the following statements is true? (Hint: Recall that $\sum_{n=1}^{\infty} 0.5^n = 1$.)

(a) $g(x_1 \ldots x_n, n) = \frac{1}{4^{n-1}}$ gives a valid language model, and there are other definitions of g that give a valid language model.

(b) $g(x_1 \ldots x_n, n) = \frac{1}{4^n}$ gives a valid language model, and there are other definitions of g that give a valid language model.

(c) $g(x_1 \ldots x_n, n) = \frac{1}{4^{n-1}}$ is the only definition of g that gives a valid language model.

(d) $g(x_1 \ldots x_n, n) = \frac{1}{4^n}$ is the only definition of g that gives a valid language model.

Part #2 8 points

For the following two questions, write TRUE or FALSE below the question. **PLEASE GIVE JUSTIFICATION FOR YOUR ANSWERS: AT MOST 50% CREDIT WILL BE GIVEN FOR ANSWERS WITH NO JUSTIFICATION.**

For both questions we assume as usual that a language model consists of a vocabulary $\mathcal{V}$, and a function $p(x_1 \ldots x_n)$ such that for all sentences $x_1 \ldots x_n \in \mathcal{V}^\dagger$, $p(x_1 \ldots x_n) \geq 0$, and in addition $\sum_{x_1 \ldots x_n \in \mathcal{V}^\dagger} p(x_1 \ldots x_n) = 1$. Here $\mathcal{V}^\dagger$ is the set of all sequences $x_1 \ldots x_n$ such that $n \geq 1$, $x_i \in \mathcal{V}$ for $i = 1 \ldots (n-1)$, and $x_n = \text{STOP}$.

We assume throughout this question that all words seen in any test corpus are in the vocabulary $\mathcal{V}$.

Question 1 (4 points) True or False? For any language model, there is at least one test corpus that has perplexity equal to ∞ under the language model.

Question 2 (4 points) True or False? (3 points): For any test corpus, there is at least one language model that has perplexity equal to ∞ on the test corpus.

Part #3 10 points

Consider the following HMM:

- Vocabulary $\mathcal{V} = \{\text{the, dog, saw, in}\}$
- Set of tags $\mathcal{K} = \{\text{D, N, P, V}\}$
- Parameters $e(\text{the}|\text{D}) = e(\text{dog}|\text{N}) = e(\text{in}|\text{P}) = e(\text{saw}|\text{V}) = 1$, with all other $e(x|s)$ parameters equal to 0.
- Parameters $q(\text{D}|*) = q(\text{N}|\text{D}) = q(\text{D}|\text{P}) = q(\text{D}|\text{V}) = 1$, and $q(\text{P}|\text{N}) = q(\text{V}|\text{N}) = q(\text{STOP}|\text{N}) = 1/3$, with all other q parameters equal to 0.

For any sentence $x_1 \ldots x_n$, we say the sentence is *grammatical under the HMM* if there is some tag sequence $y_1 \ldots y_{n+1}$ such that

$$p(x_1 \ldots x_n, y_1 \ldots y_{n+1}) > 0$$

Question 3 (5 points) Consider the following context-free grammar:

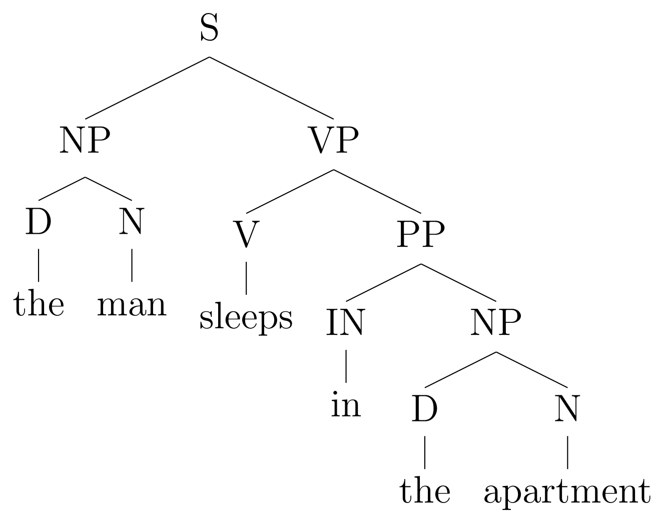
S → NP VP
NP → D N
NP → NP PP
PP → P NP
VP → V NP

Write down a sentence that is grammatical under the HMM, but is *not* grammatical under this context-free grammar.

Question 4 (5 points) Write down a context-free grammar such that the set of sentences that are grammatical under the grammar is *identical* to the set of sentences that are grammatical under the HMM given above.

Part #4 (10 points)

Consider the following parse tree:



And in addition consider the following rules that can be used to lexicalize the parse tree (note that these rules do not necessarily make sense from a linguistic perspective):

- For the rule `S -> NP VP`, we define `VP` to be the head
- For the rule `NP -> D N`, we define `D` to be the head
- For the rule `VP -> V PP`, we define `PP` to be the head
- For the rule `PP -> IN NP`, we define `IN` to be the head

Recall that for a lexicalized PCFG in Chomsky Normal form, each rule takes one of the following forms:

- $X(h) \rightarrow_1 Y_1(h)Y_2(m)$ where X, Y_1, Y_2 are non-terminals, and h, m are words
- $X(h) \rightarrow_2 Y_1(m)Y_2(h)$ where X, Y_1, Y_2 are non-terminals, and h, m are words
- $X(h) \rightarrow h$ where X is a non-terminal, and h is a word

Question 5 (10 points) If we lexicalize the above parse tree, then build a lexicalized PCFG with all rules seen in the tree, what is the complete set of rules in the grammar? (You do not need to include probabilities for the rules, just list the rules in the grammar.)

Part #5 15 points

In this question we will develop an algorithm that takes the following inputs:

- A sentence $x_1 \dots x_n$
- An HMM with parameters $q(s|u, v)$ and $e(x|s)$

and which returns *the number of tag sequences* $y_1 \dots y_{n+1}$ *such that* $p(x_1 \dots x_n, y_1 \dots y_{n+1}) > 0$. That is, the algorithm should return the number of tag sequences for $x_1 \dots x_n$ that have greater than 0 probability.

The following definitions will help:

- For any trigram of tags u, v, s, we define $f(u, v, s) = 1$ if $q(s|u, v) > 0$, and $f(u, v, s) = 0$ if $q(s|u, v) = 0$.
- For any word x and tag s, we define $g(x, s) = 1$ if $e(x|s) > 0$, and $g(x, s) = 0$ if $e(x|s) = 0$.

In addition, you may assume that for any pair of tags u, v, we have $q(\text{STOP}|u, v) > 0$.

Complete the algorithm below so that it correctly returns the number of tag sequences for $x_1 \dots x_n$ that have greater than 0 probability under the HMM (you should complete the definitions for $\pi(0, *, *)$ and $\pi(k, u, v)$):

Input: a sentence $x_1 \dots x_n$, parameters $q(s|u, v)$ and $e(x|s)$ of an HMM.
Definitions: Define $\mathcal{K}$ to be the set of possible tags. Define $\mathcal{K}_{-1} = \mathcal{K}_0 = \{*\}$, and $\mathcal{K}_k = \mathcal{K}$ for $k = 1 \dots n$.
Initialization:

Set $\pi(0, *, *) =$

Algorithm:

- For $k = 1 \dots n$,
  - For $u \in \mathcal{K}_{k-1}$, $v \in \mathcal{K}_k$,

    $\pi(k, u, v) =$

- **Return** $\sum_{u \in \mathcal{K}_{n-1}, v \in \mathcal{K}_n} \pi(n, u, v)$

Part #6 15 points

In this question we will develop an algorithm that takes the following inputs:

- A sequence of n part-of-speech tags $y_1 \ldots y_n$
- A probabilistic context-free grammar $G = (N, \Sigma, S, R, q)$. For each rule in the grammar of the form

$$X \rightarrow w$$

  where $X \in N$ is a non-terminal, and $w \in \Sigma$ is a word, we assume that X is a part-of-speech tag.

The output from the algorithm should be

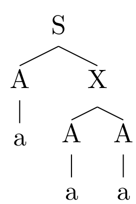
$$\max_{t \in \mathcal{T}(y_1 \ldots y_n)} p(t)$$

where $\mathcal{T}(y_1 \ldots y_n)$ is the set of all parse trees with the part-of-speech tag sequence $y_1 \ldots y_n$, and $p(t)$ is the probability of tree t under the PCFG.

As one example, given the PCFG

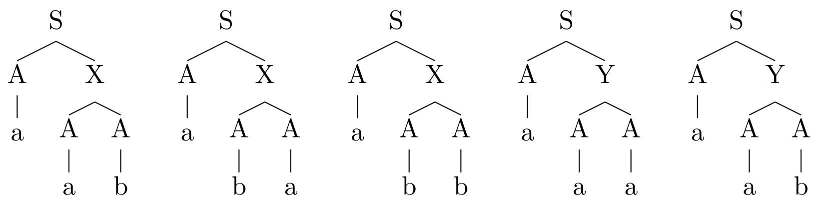
S → A X	0.6
S → A Y	0.4
X → A A	1.0
Y → A A	1.0
A → a	0.7
A → b	0.3

and the part-of-speech tag sequence $y_1 \ldots y_n$ = `A A A`, the algorithm should return the value 0.2058, because in this case the set $\mathcal{T}$(`A A A`) consists of 16 trees, of which the highest probability tree is



which has probability $0.6 \times 1.0 \times 0.7^3 = 0.2058$.

Other trees in $\mathcal{T}$(`A A A`) include



and so on.

Complete the definitions of $\pi(i, i, X)$ and $\pi(i, j, X)$ below so that the algorithm correctly returns the maximum probability for any parse tree with part-of-speech tag sequence $y_1 \dots y_n$.

Input: a sequence of part-of-speech tags $y_1 \dots y_n$, a probabilistic context-free grammar $G = (N, \Sigma, S, R, q)$.

Initialization:

For all $i \in \{1 \dots n\}$, for all $X \in N$,

$\pi(i, i, X) =$

Algorithm:

- For $l = 1 \dots (n-1)$
  - For $i = 1 \dots (n-l)$
    - Set $j = i + l$
    - For all $X \in N$, calculate

      $\pi(i, j, X) =$

Output: Return $\pi(1, n, S)$