

Ling 5801: Problem Set 4

Due via Carmen dropbox at 11:59 PM 10/31.

1. [10 pts.] Draw a recognizer chart (as in section 3 of lecture notes 10 on context-free grammar recognition) for the sentence:

‘the clown intentionally fell down twice,’

using the following grammar:

$S \rightarrow NP VP$

$NP \rightarrow \text{the clown}$

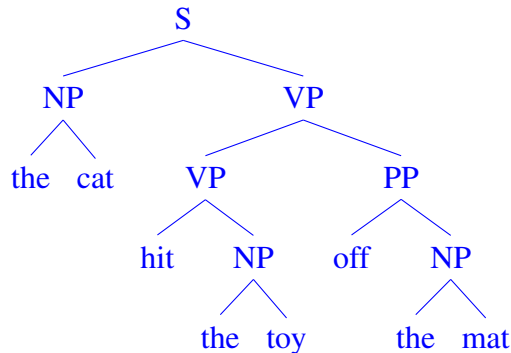
$VP \rightarrow \text{intentionally VP}$

$VP \rightarrow \text{fell down}$

$VP \rightarrow VP \text{ twice}$

2. The probability of the sentence ‘the cat hit the toy off the mat’ is defined for a probabilistic grammar in section 4 of lecture notes 11 on probabilistic parsing.

- (a) [5 pts.] Write an equation for the probability of the following tree, e.g. $P(\tau) = \dots$ in terms of grammar rule probabilities (see section 7 of lecture notes 11): $P(S \rightarrow NP VP | S)$, etc., from this grammar:



- (b) [5 pts.] Write an equation for the probability of the *sentence* ‘the cat hit the toy off the mat’: e.g. $P(X_{1..T} = \text{the cat hit the toy off the mat}) = \dots$ in terms of grammar rule probabilities: $P(S \rightarrow NP VP | S)$, etc., from this grammar. Remember that a sentence may have more than one possible phrase structure.
 - (c) [5 pts.] Give an equation for the probability of the above sentence that uses the distributive axiom to reduce the number of multiplication or addition operations in the equation.
3. Write an equation defining each of the following:
 - (a) [2 pts.] Write an equation defining a distribution $P(x \wedge y)$ for any values x and y of random variables X and Y in terms of a set of frequencies $\tilde{F}(x \wedge y)$ for all combinations of values of random variables X and Y . Your equation should be as short as possible.

- (b) [2 pts.] Write an equation defining a distribution $P(x|y)$ for any values x and y of random variables X and Y in terms of a set of frequencies $\tilde{F}(x \wedge y)$ for all combinations of values of random variables X and Y . Your equation should be as short as possible.
- (c) [3 pts.] Suppose you have counted the frequencies in some data set with which people speak slowly or quickly (above or below some threshold) when they are on or not on the telephone, and the frequencies with which they pronounce the /t/ sound (e.g. in latter) as [t] or as something like [d] when they speak quickly or slowly, on or not on the telephone. Write an equation that uses these data to calculate the estimated probability that people will pronounce /t/ as [d] given that they are on the telephone, when the speech rate is unknown.
- (d) [3 pts.] In the previous subproblem, how could you simplify this calculation if you think that the pronunciation of /t/ just depends on the speech rate, but not on whether the speaker is or is not on the telephone?
4. [10 pts.] PROGRAMMING: Write a program to read in models for variables A and B in the following format, using the `CondModel` class from the `model.py` module described in the lecture notes on probability models (and posted individually on the web site):

```
A : on = .7
A : off = .3
B on : one = .2
B on : two = .8
B off : one = .3
B off : two = .3
B off : three = .4
```

and use these to calculate $P(b)$ for each modeled value of B using the following equation:

$$P(b) = \sum_{a \in A} P(a) \cdot P(b|a)$$

5. [10 pts.] PROGRAMMING: Write a program to read in models for speaking on a phone $P(p)$, speech rate $P(r|p)$, and pronunciation of /t/ $P(t|p, r)$, in the following format:

```
P : on = .7
P : off = .3
:
R on : fast = .6
R on : slow = .4
R off : fast = .5
R off : slow = .5
:
```

```
T on fast : [t] = .9
T on fast : [d] = .1
T on slow : [t] = .8
T on slow : [d] = .2
T off fast : [t] = .3
⋮
```

and use them to calculate and print a probability model for $P(t|p)$ in the format:

```
T on : [t] = .86
T on : [d] = .14
⋮
```

Your program should be as short as possible. You may want to use the `CondModel` class from the `model.py` module described in the lecture notes on probability models (and posted individually on the web site), to help you read and write in this format. Hand in all inputs and outputs.