CS388: Midterm Exam

March 9, 2011

NAME:______________________________________

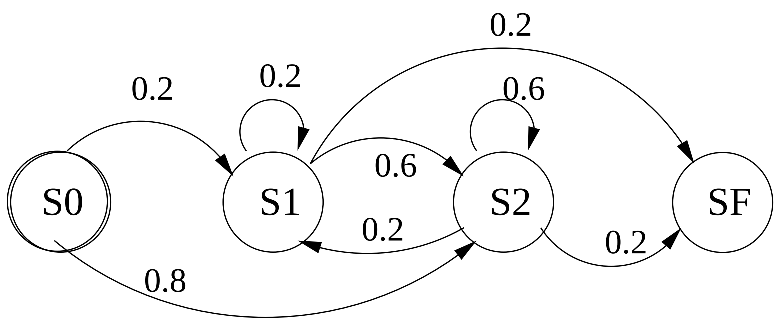
1. (20 points) Assume a bigram language model is trained on the following corpus of sentences using MLE with linear interpolation for smoothing (with the bigram λ weight set to 0.9 and the unigram λ weight set to 0.1). Since the unigram model does not need to estimate P(<s>), just completely ignore the start token when estimating the unigram model.

   - <s> dictator oppress people </s>
   - <s> army oppress people </s>
   - <s> dictator topple army </s>
   - <s> people topple army </s>
   - <s> army topple dictator </s>

   What is the estimated probability of the following test string? Show your work. You only need to calculate the parameters of the model sufficient to solve this particular problem.

   - <s> people topple dictator </s>

2. (20 points) Consider the HMM below where the transition probabilities are shown in the graph and the observation probabilities (where V={A,B}) are in the tables below each state.



b1(A) = 0.6
b1(B) = 0.4

b2(A) = 0.4
b2(B) = 0.6

Use the Veterbi algorithm to compute the most likely state sequence for the the short output string:
"A B".
Show the values computed for each of the $v_t(j)$ and $bt_t(j)$ parameters as they are computed, showing your work.

3. (15 points) Given the following context-free grammar, draw trees for all parses produced for the sentence: “Guard tests like gold.” Under each tree, briefly paraphrase in English the meaning of each parse as best you can and circle the most likely interpretation.

```
S -> NP VP, S -> VP,
NP -> Adj NP,  NP -> N, NP -> NP PP,
VP -> V, VP -> V NP, VP -> VP PP
PP -> Prep NP,
Prep -> like,
N -> guard, N -> gold, N -> tests,
V -> guard, V -> tests, V -> like,
Adj -> gold,  Adj -> guard
```

4. (21 points) Consider the following simple PCFG:

```
S -> NP VP       0.6          PropNoun -> DALLAS 0.2
S -> VP          0.4          PropNoun -> MARY   0.2
NP -> NP PP      0.4          PropNoun -> BOB    0.3
NP -> PropNoun   0.6          PropNoun -> AUSTIN 0.3
VP -> Verb       0.3          Verb -> KISS       0.5
VP -> Verb NP    0.3          Verb -> MARRY      0.5
VP -> VP PP      0.4          Prep -> IN         0.4
PP -> Prep NP    1.0          Prep -> WITH       0.6
```

Use the "Inside" CKY algorithm (the "Forward" analog for PCFGs) to find the total probability for generating the sentence:
MARRY BOB IN AUSTIN
First, below show the changes to productions that require conversion to CNF:

Next, show the triangular CKY table with each cell filled with all its constituents together with their probabilities, showing your work.

5. (24 points) Provide short answers (1-3 sentences) for each of the following questions:

What two words best characterize the presence of ambiguity in natural language?

The following is a famous dialog from the disaster-movie spoof "Airplane!":

Rumack: You'd better tell the Captain we've got to land as soon as we can. This woman has to be gotten to a hospital.
Elaine Dickinson: A hospital? What is it?
Rumack: It's a big building with patients, but that's not important right now.

Explain what *specific* type of ambiguity in language understanding makes this humorous.

Do the same for this other famous "Airplane!" dialog:

Rumack: I won't deceive you, Mr. Striker. We're running out of time.
Ted Striker: Surely there must be something you can do.
Rumack: I'm doing everything I can... and stop calling me Shirley!

What is “smoothing” and why is it critical to many areas of statistical language processing?

What is the primary difference between a generative and a discriminative probabilistic model?

Why is dynamic programming so prevelant and critical in algorithms for natural language processing?

What is the difference between maximum likelihood estimation (MLE) and maximum a posteriori (MAP) training of statistical models?

What is the easiest way to use semi-supervised learning to train a generative model?

(Extra credit, 2pts) Who was the leader of the team at IBM that built Watson and who was the UT PhD graduate who was on his team?

(Extra credit, 2pts) Who developed the first computer program for parsing natural language and at what university?