CS224N: Natural Language Processing
Spring 2008
Weekly Quiz

Quiz 4

1: In the NER task, when the the suffix feature "-field" fires, it is actually quite indicative. All else being equal, what is a word that has the suffix "field" most likely to be?

- Person name
- Organization
- Movie name
- Location
- Not a named entity

2: What is NOT true about MaxEnt models?

- Also known as log-linear models
- It is trying to maximize the entropy while respecting observed evidences
- The weights for features need not sum up to 1 but must lie in [0,1].
- None of the above

3: Which of the following is true?

- An HMM is better than a MEMM because we can easily add arbitrary features in HMM.
- The disadvantage of a discriminative model is that training and optimizing parameter weights is more expensive than for a generative model.
- A trigram language model is making a Markov assumption that the probability of a word depends on all the words that appear before it.
- All of the above
- None of the above

4: A Maximum Entropy model takes the exponential "vote" for each class and normalize them to get a probability (see lecture notes). If there is a binary classification problem, and vote(c_1) = 0.2 and vote(c_2) = -0.2, which of the following is true?

- $p(c_1)+p(c_2) = 1$
- $p(c_1) > p(c_2)$
- Both of the above
- None of the above

5: Which of the following is true?

- ◻ Character substrings and word shapes are very indicative features for the NER task.
- ◻ A CMM (a.k.a. MEMM) makes a single decision at a time, conditioned on evidence from observations and also previous decisions.
- ◻ If we apply a MaxEnt classifier (a non-sequence model) to an NER task, then features can be from the observed context but not from what labels (NER tags) they have.
- ◻ All of the above
- ◻ None of the above

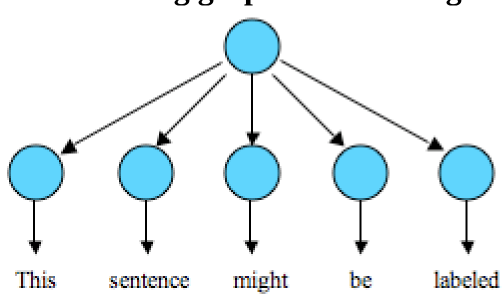
6: In a MaxEnt model, a feature C is formed as the intersection of two features A and B. If we are trying to predict many classes and using many features at once, then, when looking at the specific weights that A, B, and C have for a specific class, which of the following statements is true:

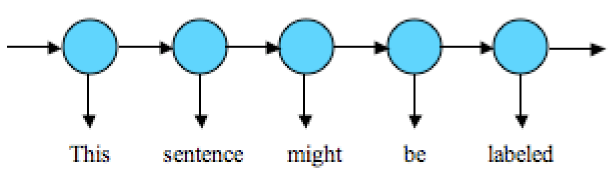
- ◻ If A and B agree in sign, C will have the same sign.
- ◻ C will have the sign of whichever A or B was greater in magnitude.
- ◻ If A and B agree in sign, C will have the opposite sign.
- ◻ C could have either sign in general.

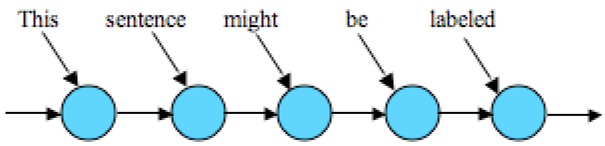
7: For an HMM parameterized by A and given a set of observations, O, the problem of finding P (A|O) is known as:

- ◻ Likelihood
- ◻ Decoding
- ◻ Learning
- ◻ Encoding
- ◻ None of the above

8: Which of the following graphs describe a generative model:

- ◻ 

- ◻ 

- 

- A & B
- A & C
- B & C

9: Which of the following does not accomplish proper regularization:

$$\arg\max_w \sum \log p(y|x) - ||w||^2$$

$$\arg\max_w \prod p(y|x) \exp\left(\frac{w^T w}{2\sigma^2}\right)$$

$$\arg\max_w \sum logp(y|x) - \sum |w|$$

$$\arg\max_w \prod \frac{p(y|x)}{\max\{1, |w|\}}$$

- a
- b
- c
- d
- All of them accomplish proper regularization

10: In MaxEnt models, we are trying to estimate the parameters w as:

$$\arg\max_w \prod_{(c,x)\in(C,X)} p(c|x,w)$$

which is equivalent to:

$$\arg\max_w \sum_{(c,x)\in(C,X)} \log p(c|x,w)$$

Recall that the final form of p(c|x,w) in a MaxEnt model is:

$$p(c|x,w) = \frac{\sum_i w_{ci} f_i(c,x)}{\sum_{c'} \exp\left(\sum_i w_{c'i} f_i(c',x)\right)}$$

Finding the optimal weights that satisfy the above equations then guarantees that the derivatives of this function are zero for all weights $w_{c\text{-hat},j}$, or:

$$\frac{\partial}{\partial w_{\hat{c}j}} \sum_{(c,x)\in(C,X)} p(c|x,w) = 0$$

This condition then guarantees that which terms are equal:

- ○ $\sum_{(c,x)\in(C,X)} f_j(c,x) = \sum_{(c,x)\in(C,X)} f_j(\hat{c},x)p(\hat{c}|x,w)$
- ○ $\sum_{\{(c,x)\in(C,X)\,:\,c=\hat{c}\}} f_j(c,x) = \sum_{(c,x)\in(C,X)} f_j(\hat{c},x)p(\hat{c}|x,w)$
- ○ $\sum_{(c,x)\in(C,X)} f_j(c,x) = \sum_{\{(c,x)\in(C,X)\,:\,c=\hat{c}\}} f_j(\hat{c},x)p(\hat{c}|x,w)$
- ○ $\sum_{\{(c,x)\in(C,X)\,:\,c=\hat{c}\}} f_j(c,x) = \sum_{\{(c,x)\in(C,X)\,:\,c=\hat{c}\}} f_j(\hat{c},x)p(\hat{c}|x,w)$
- ○ None of the above