Feedback — Quiz 1: covers material from weeks 1 and 2

Help Center

You submitted this quiz on **Sun 17 Mar 2013 10:38 AM PDT**. You got a score of **6.83** out of **7.00**. You can attempt again, if you'd like.

This is an open note quiz: you can use the slides from the class, and the notes at http://www.cs.columbia.edu/~mcollins/notes-spring2013.html as a resource.

Question 1

Say we'd like to derive the Viterbi algorithm for a **bigram** HMM tagger. The model takes the form $p(x_1 \dots x_n, y_1 \dots y_{n+1}) = \prod_{i=1}^{n+1} q(y_i|y_{i-1}) \prod_{i=1}^{n} e(x_i|y_i)$. Which of the following statements is true?

Your Answer		Score	Explanation
☑ We can implement the Viterbi algorithm in exactly the same way as before, but with the following modification to the recursive definition: $\pi(k,u,v) = \max_{w \in \mathcal{S}_{k-2}} (\pi(k-1,w,u) \times q(v|u) \times e(x_k|v))$	✔	0.33	
☐ We can use a dynamic programming algorithm with entries $\pi(k,u)$, and definitions $\pi(0,*) = 1$ and $\pi(k,v) = \max_{u \in \mathcal{S}_{k-1}} (\pi(k-2,u) \times q(v|u) \times e(x_k|v))$	✔	0.33	
☑ We can use a dynamic programming algorithm with entries $\pi(k,u)$, and definitions $\pi(0,*) = 1$ and $\pi(k,v) = \max_{u \in \mathcal{S}_{k-1}} (\pi(k-1,u) \times q(v|u) \times e(x_k|v))$	✔	0.33	
Total		1.00 / 1.00	

Question 2

Say we define a backed-off model $q_{\text{BO}}(w_i|w_{i-1})$ exactly as we defined it in lecture, and we define the discounted counts as $\text{Count}^*(w_{i-1}, w_i) = \text{Count}(w_{i-1}, w_i) - 1.5$. Which of the following statements is true?

Your Answer		Score	Explanation
☑ There may be some bigrams u, v such that $q_{\text{BO}}(v|u) < 0$	✔	0.50	
☐ There may be some words u such that $\sum_{v \in \mathcal{V} \cup \{\text{STOP}\}} q_{\text{BO}}(v|u) \neq 1$.	✔	0.50	
Total		1.00 / 1.00	

Question 3

Consider the following two bigram language models (recall that a bigram language model defines $p(x_1 \dots x_n) = \prod_{i=1}^{n} q(x_i|x_{i-1})$).

Language Model 1

$\mathcal{V} = \{\text{the}, \text{dog}\}$

$q(\text{the}|*) = q(\text{dog}|\text{the}) = q(\text{STOP}|\text{dog}) = 1$

All other q parameters are equal to 0.

Language Model 2

$\mathcal{V} = \{\text{the}, \text{a}, \text{dog}\}$

$q(\text{the}|*) = q(\text{a}|*) = 0.5$

$q(\text{dog}|\text{a}) = q(\text{dog}|\text{the}) = q(\text{STOP}|\text{dog}) = 1$

All other q parameters are equal to 0.

Now assume that we have a test sentence consisting of a single sentence,

- the dog STOP

Which language model gives **lower** perplexity on this test corpus?

Your Answer		Score	Explanation
Language Model 1	✔	1.00	
Language Model 2			
Total		1.00 / 1.00	

Question 4

We are now going to derive a version of the Viterbi algorithm that takes as input an integer n, and finds

$$\max_{y_1 \dots y_{n+1}, x_1 \dots x_n} p(x_1 \dots x_n, y_1 \dots y_{n+1})$$

for a trigram tagger, as defined in lecture. Hence the input to the algorithm is an integer n, and the output from the algorithm is the highest scoring **pair** of sequences $x_1 \dots x_n$, $y_1 \dots y_{n+1}$ under the model.

Which of the following recursive definitions gives a correct algorithm for this problem?

Your Answer		Score	Explanation
$\pi(0,*,*)=1$, and $\pi(k,u,v)=\max_{w\in\mathcal{S}_{k-2}}(\pi(k-1,w,u)\times q(v\|w,u))$			
$\pi(0,*,*)=1$, and $\pi(k,u,v)=\max_{w\in\mathcal{S}_{k-2}}(\pi(k-1,w,u)\times q(v\|w,u)\times m(v))$, where $m(v)=\max_{x\in\mathcal{V}} e(x\|v)$	✔	1.00	
None of the above.			
Total		1.00 / 1.00	

Question 5

We'd like to define a language model with $\mathcal{V}=\{\mathrm{the},\mathrm{a},\mathrm{dog}\}$, and

$$p(x_1\dots x_n)=\gamma\times 0.5^n$$

for any $x_1\dots x_n$, such that $x_i\in\mathcal{V}$ for $i=1\dots(n-1)$, and $x_n=\mathrm{STOP}$, where γ is some expression.

What should our definition of γ be?

(Hint: recall that $\sum_{n=1}^{\infty}0.5^n=1$)

Your Answer	Score	Explanation
$\gamma=\frac{1}{3^n}$		

$\gamma = 3^n$		
$\gamma = 3^{n-1}$		
$\gamma = 1$		
$\gamma = \frac{1}{3^{n-1}}$	✔	1.00
Total		1.00 / 1.00

Question 6

Say we train a trigram HMM tagger on a training set with the following two sentences:

- the dog saw the cat, D N V D N
- the cat saw the saw, D N V D N

Assume that we estimate the parameters of the HMM with maximum-likelihood estimation (no smoothing).

Now assume that we have the sentence

$x_1 \ldots x_n =$ the cat saw the saw

what is the value for

$\max_{y_1 \ldots y_{n+1}} p(x_1 \ldots x_n, y_1 \ldots y_{n+1})$ in this case? (Please give your answer up to 3 decimal places.)

You entered:

0.031

Your Answer		Score	Explanation
0.031	✔	1.00	
Total		1.00 / 1.00	

Question 7

Assume we have a bigram language model with

$\mathcal{V} = \{\mathrm{the}, \mathrm{a}\}$

$q(\mathrm{a}|*) = 0.6$, $q(\mathrm{the}|*) = 0.4$, $q(\mathrm{a}|\mathrm{a}) = 0.9$, $q(\mathrm{STOP}|\mathrm{a}) = 0.1$, $q(\mathrm{the}|\mathrm{the}) = 0.8$, $q(\mathrm{STOP}|\mathrm{the}) = 0.2$, all other parameter values equal to 0.

Now say we'd like to define a bigram HMM model which defines the same distribution over sentences as the language model. By this we mean the following. The bigram HMM defines a distribution over sentences $x_1 \ldots x_n$ paired with tag sequences $y_1 \ldots y_{n+1}$ as follows:

$p'(x_1 \ldots x_n, y_1 \ldots y_{n+1}) = \prod_{i=1}^{n+1} q'(y_i|y_{i-1}) \prod_{i=1}^{n} e'(x_i|y_i)$

(Note we use the notation p', q' and e' to distinguish this from the distribution p and parameters q in the language model.)

The bigram HMM defines the same distribution over sentences as the language model if for any sentence $x_1 \ldots x_n$,

$$p(x_1 \ldots x_n) = \sum_{y_1 \ldots y_{n+1}} p'(x_1 \ldots x_n, y_1 \ldots y_{n+1})$$

where p and p' are the distributions under the language model and the bigram HMM respectively.

Our HMM will have a set of tags $\mathcal{S} = \{1, 2\}$, and a vocabulary $\mathcal{V} = \{\text{the}, \text{a}\}$. We define $q'(1|*) = 0.6$.

In this question you should choose the parameters of the HMM so that it gives the same distribution over sentences as the language model given above. What should be the values for $q'(2|*)$, $q'(1|1)$, $q'(2|1)$, $q'(\text{STOP}|1)$, $e'(\text{the}|1)$, $e'(\text{the}|2)$?

Write your answers in order in the box below, separated by spaces. For example, you could write

0.2 0.3 1 0 0.4 0.5

You entered:

0.4 0.9 0.1 0.1 0 1

Your Answer		Score	Explanation
0.4	✔	0.17	
0.9	✔	0.17	
0.1	✖	0.00	
0.1	✔	0.17	

0	✔	0.17
1	✔	0.17
Total		0.83 / 1.00