ECS763U/ECS763P - Natural Language Processing - 2022/23 - Semester 1

Started on	Monday, 24 October 2022, 10:23 PM
State	Finished
Completed on	Monday, 24 October 2022, 10:52 PM
Time taken	28 mins 23 secs
Grade	**10.00** out of 10.00 (**100**%)

Quiz navigation

1 2 3 4 5 6 7 8 9 10

Show one page at a time

Finish review

Question **1**
Correct
Mark 1.00 out of 1.00
Flag question

A better model of a text is one which assigns a _____ probability to the words that actually occur in the corpus.

Select one:
- a. higher ✓
- b. lower

The correct answer is: higher

Question **2**
Correct
Mark 1.00 out of 1.00
Flag question

For any amount of natural language data, unless this data only consists of one word repeating itself over and over again there will always be more bigram types than unigram types.

Select one:
- True
- False ✓

The correct answer is 'False'.

Question **3**
Correct
Mark 1.00 out of 1.00
Flag question

It generally doesn't matter if you use different vocabularies when comparing different language models against each other in terms of perplexity.

Select one:
- True
- False ✓

The correct answer is 'False'.

Question **4**
Correct
Mark 1.00 out of 1.00
Flag question

A better model of a text is one which assigns a _____ perplexity to the words that actually occur in the corpus.

Select one:
- a. lower ✓
- b. higher

The correct answer is: lower

Question **5**
Correct
Mark 1.00 out of 1.00
Flag question

Which of these are smoothing techniques in language modelling?

Select one or more:
- a. Interpolation with lower order models. ✓
- b. Spelling normalisation.
- c. Feature reduction.
- d. Tokenisation.
- e. Absolute discounting. ✓
- f. Laplace. ✓

The correct answers are: Absolute discounting., Laplace., Interpolation with lower order models.

Question **6**
Correct
Mark 1.00 out of 1.00
Flag question

In creating the counts for a bigram language model, the word `John' has been counted 4 times in a corpus, the bigram `John runs' has been counted 3 times and a total of 6 different word types have been observed in the corpus overall. What is the Laplace (add-one) smoothed probability of $p(runs|John)$?

Select one:
- a. 0.4 ✓
- b. 0.8
- c. 0.666 to 3 d.p.
- d. 0.5
- e. 0.556 to 3 d.p.

The correct answer is: 0.4

Question **7**
Correct
Mark 1.00 out of 1.00
Flag question

Which of the following is true?

Select one:
- a. A Kneser-Ney smoothed model uses the counts of n-gram types for all its n-gram models.
- b. A Kneser-Ney smoothed model uses the counts of n-gram tokens for its highest order n-gram model, but the counts of n-gram types for its lower order n-gram models which it interpolates with. ✓
- c. A Kneser-Ney smoothed model uses the counts of n-gram tokens for all its n-gram models.
- d. A Kneser-Ney smoothed model does not use the counts of n-gram tokens for any of its n-gram models.
- e. A Kneser-Ney smoothed model uses the counts of n-gram types of its highest order n-gram model, but the counts of n-gram tokens for its lower order n-gram models which it interpolates with.

The correct answer is: A Kneser-Ney smoothed model uses the counts of n-gram tokens for its highest order n-gram model, but the counts of n-gram types for its lower order n-gram models which it interpolates with.

Question **8**
Correct
Mark 1.00 out of 1.00
Flag question

The generalisation of the product rule of probability theory for more than two variables is called the chain ✓ rule of probability.

The correct answer is: chain

Question **9**
Correct
Mark 1.00 out of 1.00
Flag question

Consider the training data below for a language model consisting of three sentences. Note the padding around the words- you should consider both the beginning and end of sentence markers as words.

To obtain the counts on the sentences each one will be scanned through and each word counted as per the two functions C in the formula for a bigram model shown below. Remember that in the count in the denominator the word $\langle/s\rangle$ for the end of sentence will never be counted. Note that the size of vocabulary V will not include the beginning padding marker $\langle s\rangle$ but can contain the end of sentence marker $\langle/s\rangle$. After the counts have been collected, according to a Laplace (add-one) smoothed bigram model, what is the raw (i.e. not log) probability value for $p(John|likes)$. Give your answer to 1 DECIMAL PLACE.

$$p(w_i|w_{i-1}) = \frac{C(w_{i-1}, w_i) + 1}{C(w_{i-1}) + V}$$

(note beginning <s>) and end-of-sentence (</s>) markers, treat them as words):

<s> John likes Mary and Bill </s>
<s> Mary likes John and Mohammed </s>
<s> Mary and John like Mohammed and Bill </s>

Answer: 0.2 ✓

The correct answer is: 0.2

Question **10**
Correct
Mark 1.00 out of 1.00
Flag question

Consider the training data below for a language model consisting of three sentences. Note the padding around the words- you should consider both the beginning and end of sentence markers as words.

To obtain the counts on the sentences each one will be scanned through and each word counted as per the two functions C in the formula for a bigram language model shown below. Remember that in the count in the denominator the word $\langle/s\rangle$ for the end of sentence will never be counted. After the counts have been collected, according to a maximum likelihood estimation bigram model, what is the raw (i.e. not log) probability value for $p(John|likes)$. Give your answer to 1 DECIMAL PLACE.

$$p(w_i|w_{i-1}) = \frac{C(w_{i-1}, w_i)}{C(w_{i-1})}$$

(note beginning <s>) and end-of-sentence (</s>) markers, treat them as words):

<s> John likes Mary and Bill </s>
<s> Mary likes John and Mohammed </s>
<s> Mary and John like Mohammed and Bill </s>

Answer: 0.5 ✓

The correct answer is: 0.5

Finish review