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

Started on	Monday, 24 October 2022, 10:56 AM
State	Finished
Completed on	Monday, 24 October 2022, 11:21 AM
Time taken	24 mins 50 secs
Grade	**7.00** out of 10.00 (**70**%)

Quiz navigation

1 2 3 4 5 6 7 8 9 10

Show one page at a time

Finish review

Question **1**
Incorrect
Mark 0.00 out of 1.00
Flag question

It's always the case that a higher number for n in n-grams gives a lower perplexity score.

Select one:
- True ✗
- False

The correct answer is 'False'.

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

Language models can be used as part of speech recognition.

Select one:
- True ✓
- False

The correct answer is 'True'.

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

It's generally the case for a large amount of natural language data that a bigram model will give a lower perplexity score than a unigram model with the same vocabulary.

Select one:
- True ✓
- False

The correct answer is 'True'.

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

Language models can be used as part of machine translation.

Select one:
- True ✓
- False

The correct answer is 'True'.

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

Which of these are applications which use language models?

Select one or more:
- a. Sentiment analysis only using the binary feature of whether a word is in the text or not.
- b. Sentiment analysis only using the binary feature of whether a bigram is in the text or not.
- c. Sentiment analysis only using the perplexity of a model of relative word frequency in a given class. ✓
- d. Automatic speech recognition (ASR). ✓
- e. Finding a keyword in a text.
- f. Machine translation. ✓

The correct answers are: Automatic speech recognition (ASR)., Machine translation., Sentiment analysis only using the perplexity of a model of relative word frequency in a given class.

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

In creating the counts for a bigram language model, the word ' Mohammed' has been counted 7 times in a corpus, the bigram ' Mohammed runs' has been counted 3 times, and a total of 15 different word types have been observed in the corpus overall. What is the Laplace (add-one) smoothed probability of $p(runs|Mohammed)$?

Select one:
- a. 0.429 to 3 d.p.
- b. 0.571 to 3 d.p.
- c. 0.15
- d. 0.2 ✓
- e. 0.5

The correct answer is: 0.2

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

Which of the following should be done first in building an n-gram language model?

Select one:
- a. Get the probabilities for the model.
- b. Define n (the order) of the model. ✓
- c. Smooth the model.
- d. Get the frequency counts for the different ngrams.
- e. Define the discount of the model.

The correct answer is: Define n (the order) of the model.

Question **8**
Incorrect
Mark 0.00 out of 1.00
Flag question

The most common method for comparing the performance of language models is computing the cross-entropy ✗ of the model on a corpus, where the lower the result, the better.

The correct answer is: perplexity

Question **9**
Incorrect
Mark 0.00 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 $p(Mary|likes)$. Give your answer to 1 DECIMAL PLACE.

$$p(w_j \mid w_{j-1}) = \frac{C(w_{j-1}, w_j) + 1}{C(w_{j-1}) + V}$$

(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: 0.1 ✗

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 $p(John|and)$. Give your answer to 2 DECIMAL PLACES.

$$p(w_j \mid w_{j-1}) = \frac{C(w_{j-1}, w_j)}{C(w_{j-1})}$$

(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: 0.25 ✓

The correct answer is: 0.25

Finish review

◀ Week 3 Lecture Zoom Link | Jump to... | Unit 4 Lecture Slides ▶