|  |  |
| --- | --- |
| **Started on** | Wednesday, March 24, 2021, 2:01 PM |
| **State** | Finished |
| **Completed on** | Sunday, March 28, 2021, 9:56 PM |
| **Time taken** | 4 days 7 hours |
| **Points** | 7.00/8.00 |
| **Grade** | **87.50** out of 100.00 |

Top of Form

Question **1**

Correct

1.00 points out of 1.00

Flag question

Question text

**If we trained a trigram model and a bigram model on the same very large dataset, we would expect the perplexity of the trigram model to be \_\_\_\_\_ than the bigram model:**

Select one:

a. The same as

b. We can't predict because of different vocabulary sizes

c. Lower

d. Higher

Feedback

The correct answer is: Lower

Question **2**

Correct

1.00 points out of 1.00

Flag question

Question text

Rule-based NLP is:

Select one:

a. **Costly and burdensome compared to machine learning**

b. **Widely used and sought-after**

c. **Based on simple probabilities**

d. **Known to lower the perplexity of your language models**

Feedback

The correct answer is: **Costly and burdensome compared to machine learning**

Question **3**

Correct

1.00 points out of 1.00

Flag question

Question text

**NLP skills are \_\_\_\_ and \_\_\_\_\_:**

Select one:

a. **In-demand, something most data science professionals have mastered**

b. **In-demand, rarely possessed by data science professionals**

c. **Specialized and niche, rarely possessed by data science professionals**

d. **Specialized and niche, something most data science professionals have mastered**

Feedback

The correct answer is: **In-demand, rarely possessed by data science professionals**

Question **4**

Correct

1.00 points out of 1.00

Flag question

Question text

**According to the Markov assumption, the probability of the word “she” given the sequence “my pug is so cute that”, is roughly equivalent to what probability?**

Select one:

a. **The probability of the word “that” given “my pug is so cute”**

b. **The probability of the word “she” given the sequence “pug is so cute that”**

c. All of the above

d. **The probability of the word “she” given “that”**

Feedback

The correct answer is: All of the above

Question **5**

Incorrect

0.00 points out of 1.00

Flag question

Question text

**Which of the following is not a problem that makes NLP hard compared to other data science tasks?**

Select one:

a. **All of the above are problems that make NLP hard compared to other data science tasks**

b. **Words cannot be converted into numbers**

c. **The inputs of models are not always the same length**

d. **Words are not numbers**

Feedback

The correct answer is: **Words cannot be converted into numbers**

Question **6**

Correct

1.00 points out of 1.00

Flag question

Question text

**Which task is well-suited to creating word vectors?**

Select one:

a. Bigrams

b. Classification

c. Language modeling

d. Sentiment analysis

Feedback

The correct answer is: Language modeling

Question **7**

Correct

1.00 points out of 1.00

Flag question

Question text

**The purpose of smoothing is to:**

Select one:

a. **Make sure that your language model is trained on an adequately large dataset**

b. **Make sure you don’t repeat <unk> ad nauseum**

c. **Make sure that unattested words in your dataset don’t cause your probability to equal zero**

d. **Make sure that your bigram model takes into account surrounding context so you don’t get ungrammatical collocations**

Feedback

The correct answer is: **Make sure that unattested words in your dataset don’t cause your probability to equal zero**

Question **8**

Correct

1.00 points out of 1.00

Flag question

Question text

**In order to avoid floating point underflow in your code when implementing ngram language models, ensure that you:**

Select one:

a. **Log and sum your probabilities rather than multiplying them**

b. **Never have zeros in your numerators**

c. **Always use smoothing**

d.  
**Log your probabilities before multiplying them**

Feedback

The correct answer is: **Log and sum your probabilities rather than multiplying them**

Bottom of Form