1. Corpus: “In every place of great resort the monster was the fashion. They sang of it in the cafes, ridiculed it in the papers, and represented it on the stage. ” (Jules Verne, Twenty Thousand Leagues under the Sea)

1 / 1 point

In the context of our corpus, what is the probability of word “papers” following the phrase “it in the”.

- ○ P(papers|it in the) = 0
- ○ P(papers|it in the) =1
- ○ P(papers|it in the) = 2/3
- ◉ P(papers|it in the) = 1/2

Correct
Correct

2. Given these conditional probabilities

1 / 1 point

P(Mary)=0.1; P(likes)=0.2; P(cats)=0.3 . P(Mary|likes) =0.2; P(likes|Mary) =0.3; P(cats|likes)=0.1; P(likes|cats)=0.4

Approximate the probability of the following sentence with bigrams: “Mary likes cats”

- (●) P(Mary likes cats) = 0.003
- ( ) P(Mary likes cats) = 0.008
- ( ) P(Mary likes cats) = 0
- ( ) P(Mary likes cats) =1

Correct
Correct.

3. Given these conditional probabilities

1 / 1 point

P(Mary)=0.1; P(likes)=0.2; P(cats)=0.3

P(Mary|<s>)=0.2; P(</s>|cats)=0.6

P(likes|Mary) =0.3; P(cats|likes)=0.1

Approximate the probability of the following sentence with bigrams: "<s> Mary likes cats </s>"

- ○ P(<s> Mary likes cats </s>) = 0.003
- ○ P(<s> Mary likes cats </s>) = 1
- ○ P(<s> Mary likes cats </s>) = 0
- ◉ P(<s> Mary likes cats </s>) = 0.0036

✓ **Correct**

4. Given the logarithm of these conditional probabilities: **1 / 1 point**

log(P(Mary|<s>))=-2; log(P(</s>|cats))=-1

log(P(likes|Mary)) =-10; log(P(cats|likes))=-100

Approximate the log probability of the following sentence with bigrams : “<s> Mary likes cats </s>”

- (●) log(P(<s> Mary likes cats </s>)) = -113
- ( ) log(P(<s> Mary likes cats </s>)) = 2000
- ( ) log(P(<s> Mary likes cats </s>)) = -112
- ( ) log(P(<s> Mary likes cats </s>)) = 113

Correct
Correct

5. Given the logarithm of these conditional probabilities: **1 / 1 point**

log(P(Mary|<s>))=-2; log(P(</s>|cats))=-1

log(P(likes|Mary)) =-10; log(P(cats|likes))=-100

Assuming our test set is W="<s> Mary likes cats </s>", what is the model's perplexity.

- ○ log PP(W) = -113
- ○ log PP(W) = (-1/5)*(-113)
- ◉ log PP(W) = (-1/4)*(-113)
- ○ log PP(W) = (-1/5)*113

Correct
Correct.

6. Given the training corpus and minimum word frequency=2, how would the vocabulary for corpus preprocessed with <UNK> look like? **1 / 1 point**

"<s> I am happy I am learning </s> <s> I am happy I can study </s>"

- ○ V = (I,am,happy,learning,can,study)
- ○ V = (I,am,happy,learning,can,study,<UNK>)
- ○ V = (I,am,happy,I,am)
- ◉ V = (I,am,happy)

Correct
Correct

7. Corpus: “I am happy I am learning”

1 / 1 point

In the context of our corpus, what is the estimated probability of word “can” following the word “I” using the bigram model and add-k-smoothing where k=3.

- ○ P(can|I) = 0
- ○ P(can|I) =1
- ◉ P(can|I) = 3/(2+3*4)
- ○ P(can|I) = 3/(3*4)

Correct

Correct.

8. Which of the following are applications of n-gram language models? **1 / 1 point**

- [x] Speech recognitions

  Correct
  Correct

- [x] Auto-complete

  Correct
  Correct

- [x] Auto-correct

  Correct
  Correct

- [x] Augmentative communication

  Correct
  Correct

- [ ] Sentiment Analysis

9. The higher the perplexity score the more our corpus will make sense. **1 / 1 point**

- ○ True
- ◉ False

Correct
Correct.

10. The perplexity score increases as we increase the number of <UNK> tokens. **0 / 1 point**

- (●) True.
- ( ) False.

Incorrect
Correct

10. The perplexity score increases as we increase the number of <UNK> tokens. **1 / 1 point**

- ○ True.
- ◉ False.

Correct
Incorrect.