Assignment 2

Unit 4 - Week 2

1. Consider the following context C, a 4-sentence story. What is the total count of unique nouns in this story?

   a. Count: 20
   b. 21
   c. 22
   d. 23

2. A 4-gram model is a ________ of the Markov Model.
   a. Constraint
   b. First
   c. Fourth
   d. None

3. Average the sentence: "Teacher, teacher, teacher, teacher."

4. For the string "Hello world", identify which of the following are not parts of the language: a. small, big, book, book
   b. hash, hash, hash, basis
   d. None of the above

5. Assume that we re-use the score we derived for operation S in calculating Levenshtein distance, such that for the triangle and diagonal operations have a cost of -1 each, while substitutions costs 2 each. What is the total score for the operation "Hello world" which differs from the string "Hello world"? What is the minimum number of operations required to transform one string to another?

6. Calculate PPL (play a big game) meaning a big language model.

7. Everything the same except in question (b), calculate the probability of the word "play a big game".

8. Assume that we are using an 8-gram language model with zero smoothing. Calculate PPL (play a big game) meaning a big language model.

9. Which of the following sentences will be most likely to occur according to the language model of question (b)?

   a. Our vocabulary is limited
   b. Our vocabulary is limited
   c. Our vocabulary is limited
   d. A writer is a writer.