Introduction to Large Language Models

Week-2 Assignment

Number of questions: 8 **Total mark: 6 X 1 + 2 x 2 = 10**

Question 1:

Which of the following does **not** directly affect perplexity?

a. Vocabulary size
b. Sentence probability
c. Number of tokens
d. Sentence length

Answer: a

Question 2:

What is the goal of a probabilistic language model?

a. Translate sentences
b. Predict the next word in a sequence
c. Classify documents
d. Summarize text

Answer: b

Question 3:

Which equation expresses the chain rule for a 4-word sentence?
a. $P(w1, w2, w3, w4) = P(w1) + P(w2|w1) + P(w3|w2) + P(w4|w3)$
b. $P(w1, w2, w3, w4) = P(w1) \times P(w2|w1) \times P(w3|w1, w2) \times P(w4|w1, w2, w3)$
c. $P(w1, w2, w3, w4) = P(w1) \times P(w2|w1) \times P(w3|w2) \times P(w4|w3)$
d. $P(w1, w2, w3, w4) = P(w4|w3) \times P(w3|w2) \times P(w2|w1) \times P(w1)$

Answer: b

Question 4:

Which assumption allows n-gram models to reduce computation?

a. Bayes Assumption
b. Chain Rule
c. Independence Assumption
d. Markov Assumption

Answer: d

Question 5:

In a trigram language model, which of the following is a correct example of linear interpolation?

a. $P(w_i|w_{i-2},w_{i-1})=\lambda_1 P(w_i|w_{i-2},w_{i-1})$
b. $P(w_i|w_{i-2},w_{i-1})=\lambda_1 P(w_i|w_{i-2},w_{i-1})+\lambda_2 P(w_i|w_{i-1})+\lambda_3 P(w_i)$
c. $P(w_i|w_{i-2},w_{i-1})=\max(P(w_i|w_{i-2},w_{i-1}),P(w_i|w_{i-1}))$

d. $P(w_i|w_{i-2},w_{i-1})=P(w_i)P(w_{i-1})/P(w_{i-2})$

Answer: b

Question 5:
A trigram model is equivalent to which order Markov model?

a. 3
b. 2
c. 1
d. 4

Answer: b

Question 6:

Which smoothing technique leverages the number of unique contexts a word appears in?

a. Good-Turing
b. Add-k
c. Kneser-Ney
d. Absolute Discounting

Correct Answer: c
Explanation: Kneser-Ney uses continuation probability which counts the number of unique left contexts.

For Question 4 to 5, consider the following corpus:
<s> the sky is blue </s>
<s> birds fly in the sky </s>
<s> the blue birds sing </s>

QUESTION 4:
Assuming a bi-gram language model, calculate the probability of the sentence:
<s> birds fly in the blue sky </s>
Ignore the unigram probability of P(<s>) in your calculation.
a. 2/37
b. 1/27
c. 0
d. 1/36
Correct Answer: c
Solution:

P(<s> birds fly in the blue sky </s>) =
P(<s>) x P(birds | <s>) × P(fly | birds) × P(in | fly) × P(the | in) × P(blue | the) × P(sky | blue) × P(</s> | sky)

From the corpus:

P(birds | <s>) = Count(<s> birds) / Count(<s>) = 1 / 3

P(fly | birds) = Count(birds fly) / Count(birds) = 1 / 2

P(in | fly) = Count(fly in) / Count(fly) = 1 / 1

P(the | in) = Count(in the) / Count(in) = 1 / 1

P(blue | the) = Count(the blue) / Count(the) = 1 / 3

P(sky | blue) = Count(blue sky) = 0

P(<s> birds fly in the blue sky </s>) = 0

QUESTION 5:
Assuming a bi-gram language model, calculate the perplexity of the sentence:
<s> birds fly in the blue sky </s>

Please do not consider <s> and </s> as words of the sentence.
a. 271/4
b. 271/5
c. 91/6
d. None of these

Correct Answer: d
Solution:
As calculated in the previous question,
P(<s> birds fly in the blue sky </s>) = 0
Thus, Perplexity = undefined
