✖ **Try again once you are ready.**

Required to pass: 60% or higher

You can retake this quiz up to 3 times every 8 hours.

Back to Week 2

Retake

1 / 1 point

1.
Given the corpus of three sentences

This is the house that Jack built.

This is the malt that lay in the house that Jack built.

This is the rat that ate the malt that lay in the house that Jack built.

calculate the probability p (house | the) using maximum likelihood estimation.

○ 1/3

◉ 1/2

Correct
This is correct. There are six occurrences "the" in the corpus and only three of them are followed by "house".

○ 1/6

○ 2

---

2 / 2 points

2.
Consider the **bigram language model** trained on the sentence:

This is the cow with the crumpled horn that tossed the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built.

Find the **probability of the sentence:**

This is the rat that worried the dog that Jack built.

○ 1/8

◉ $\frac{1}{6} \cdot \frac{1}{7} \cdot \frac{1}{6} \cdot \frac{1}{7}$

Correct
Exactly! Most of the conditional probabilities are equal to 1, e.g. p(is|This) = 1 since "This" occurs only once in the training data and it's followed by "is". Only the probabilities for "the" and "that" are non-trivial.

○ ∞

○ 0

○ $\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{7} \cdot \frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{7} \cdot \frac{1}{3} \cdot \frac{1}{5} \cdot \frac{1}{4}$

3.

Consider the **trigram language model** trained on the sentence:

This is the rat that ate the malt that lay in the house that Jack built.

Find the **perplexity** of this model on the test sentence:

This is the house that Jack built.

○   ∞

Correct
Yes. The probability p (house | is the) is zero.

○   1

○ $\dfrac{1}{\sqrt[7]{\frac{1}{3}\cdot\frac{1}{3}}} = \sqrt[7]{9}$

○   0

✖   0 / 4 points

4.

Apply **add-one smoothing** to the trigram language model trained on the sentence:

This is the rat that ate the malt that lay in the house that Jack built.

Find the **perplexity** of this smoothed model on the test sentence:

This is the house that Jack built.

Write the answer with precision of 3 digits after the decimal point.

> 9.377

Incorrect Response
Some hints for you: there are 12 unique words in train, so V=13 (because of the fake end token). The length of the test sentence is N=7 words. Use these values for the vocabulary size in add-one smoothing and for the root index in the perplexity respectively. And do not forget about start and end tokens!

You might need a piece of paper to calculate this. Get back to our reading material in this module to review a similar task.

✖   0 / 2 points

5.

Find one incorrect statement below:

○   Trigram language models can have a larger perplexity than bigram language models.

This should not be selected
The perplexity of higher order language models without smoothing may by infinite!

○   N-gram language models cannot capture distant contexts.

○   If a test corpus does not have out-of-vocabulary words, smoothing is not needed.

○ The smaller holdout perplexity is - the better the model.

○ End-of-sentence tokens are necessary for modelling probabilities of sentences of different lengths.