Tania Soutonglang

CS 585-02

Sept. 12, 2023

<div align="center">Homework 1 Problems 6-7</div>

6. Refer to your results from Problem 5. Which distribution gives you the lowest entropy rate for your movie review? Does this match what you expected? Why or why not?

Entropy rate with SST: 4.836273055258784e-07

Entropy rate with QNLI: 4.5657562774588297e-07

The QNLI distribution gave the lowest entropy rate for my movie review. I expected this only because the probability distribution of this dataset had lower numbers compared to the SST dataset, therefore anything calculated with that probability distribution would be lower.

7. Problem 5 required that you handle "zero probabilities" cases, where a token occurred in one dataset but not the other. How did you handle these tokens? (Hint: Dropping the word from both probability distributions is not an ideal solution).

I handled these cases by making the probability of the missing token 0.0. This allowed me to still calculate the word if it appeared in one dataset, but also factored in the fact that it does not exist in the other (therefore the 0% possibility in the probability distribution).