MT Summer Term 2021 Ex5: Language Model Basics

Q1: Discrete Count-Based Language Models and Perplexity

(i) Given a string of words $w_1 w_2 w_3 \ldots w_n$ what do language models (LMs) tell us?

1. …
2. …

(ii) How can you model $P(w_1 w_2 w_3 \ldots w_n)$ faithfully using the chain rule from Probability Theory? Is this modelling "faithful", i.e. do you get an exact equality?

(iii) How can you estimate the component probabilities of the chain rule in (ii) above using MLE and counts from data?

(iv) What's "wrong" (in the sense of making this not very useful) with this application of the chain rule for practical language modelling?

(v) What is the Markov assumption and how can you use the Markov assumption to overcome the practical modelling problem above? In what sense is the Markov assumption an approximation?

(vi) Please turn $P(w_1 w_2 w_3 \ldots w_n)$ into bigram model using Markov assumption and beginning and end of sentence markers $\langle s \rangle$ and $\langle /s \rangle$.

(vii) Please define surprisal and surprisal in context using probabilities and explain the intuitions.

(viii) In your own words, what is (are) the intuition(s) behind perplexity? What is perplexity used for?

$$PPL(w_1 w_2 \ldots w_n) = 2^{-\frac{\log_2 P(w_1 w_2 \ldots w_n)}{n}} = \left(P(w_1 w_2 \ldots w_n)\right)^{-\frac{1}{n}} = \sqrt[n]{\frac{1}{P(w_1 w_2 \ldots w_n)}} \ ?$$

(ix) Given a unigram LM:

$$P(x) = \frac{1}{4},\ P(y) = \frac{1}{2},\ P(z) = \frac{1}{4}$$

What is the perplexity of the string "$x\ y\ z$"?

(x) Given another unigram LM:

$$P(x) = \frac{1}{3},\ P(y) = \frac{1}{3},\ P(z) = \frac{1}{3}$$

What is the perplexity of the string “$x\ y\ z$” now?

(xi) Given the ten numerals (the digits) $0,1,2,3,4,5, \ldots, 9$, and assuming that all are equally probable, what is the perplexity of any 3 digit string?