N-Gram Language Model:

* Estimate probability of word sequences
* Estimate probability of a word following sequence of words

Application:

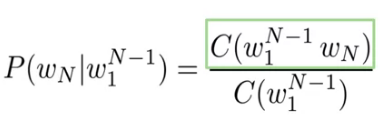
* Autocomplete a sentence with most likely suggestions
* Speech Recognition to convert the output to a real word
* Spelling correction
* Augmentative communication

Sentence Auto-complete Model:

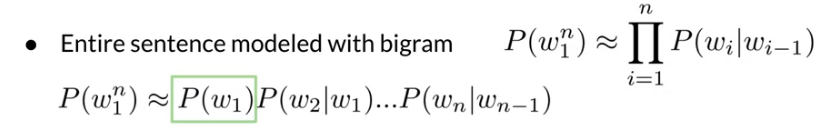
* Process text corpus to N-gram Language Model
* Out of vocabulary words
* Smoothing for previously unseen N-grams
* Language Model evaluation with perplexity metric

N-grams and Probabilities:

* An n-gram is a sequence of N words
* Unigrams are a set of all unique single words appearing in the text
* Bigrams are all sets of two words that appear side by side in the Corpus
* Trigrams represent unique triplets of words that appear in the sequence together in the Corpus.
* N-gram probabilities:



Sequence Probabilities:

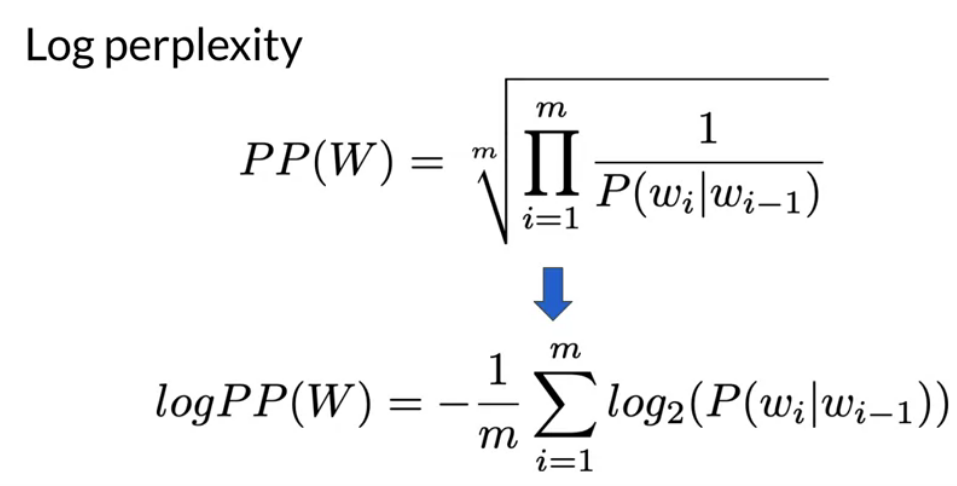


Starting and Ending Sentences:

* N-gram model: add N-1 start tokens < s >
* N-gram model: just add one < /s > to the end of the sentence

Perplexity:

The smaller the perplexity score the more likely the sentence is to sound natural to human ears. For context, good language models have perplexity scores between 60 to 20 sometimes even lower for English. Perplexities for character level language models where you track characters instead of words will be lower



* In a good model with perplexity between 20 and 60, log perplexity would be between 4.3 and 5.9

Smoothing:

* Add-k smoothing was described as a method for smoothing of the probabilities for previously unseen n-grams
* The downside is that n-grams not previously seen in the training dataset get too high probability

Back-off:

* Back-off is a model generalization method that leverages information from lower order n-grams in case information about the high order n-grams is missing. For example, if the probability of an trigram is missing, use bigram information and so on
* lambda\_factor = 0.4 is experimentally found for web-scale corpuses when using the "stupid" back-off

Interpolation:

* The other method for using probabilities of lower order n-grams is the interpolation. In this case, you use weighted probabilities of n-grams of all orders every time, not just when high order information is missing.