

≡ Item Navigation

Starting and Ending Sentences

We usually start and end a sentence with the following tokens respectively: $\langle s \rangle$ $\langle /s \rangle$.

When computing probabilities using a unigram, you can append an $\langle s \rangle$ in the beginning of the sentence. To generalize to an N-gram language model, you can add N-1 start tokens $\langle s \rangle$.

For the end of sentence token $\langle /s \rangle$, you only need one even if it is an N-gram. Here is an example:

Example - bigram

Corpus

$\langle s \rangle$ Lyn drinks chocolate $\langle /s \rangle$

$\langle s \rangle$ John drinks tea $\langle /s \rangle$

$\langle s \rangle$ Lyn eats chocolate $\langle /s \rangle$

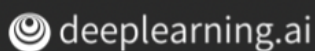
$$P(\text{sentence}) = \frac{2}{3} * \frac{1}{2} * \frac{1}{2} * \frac{2}{2} = \frac{1}{6}$$

$$P(\text{John} | \langle s \rangle) = \frac{1}{3}$$

$$P(\langle /s \rangle | \text{tea}) = \frac{1}{1}$$

$$P(\text{chocolate} | \text{eats}) = \frac{1}{1}$$

$$P(\text{Lyn} | \langle s \rangle) = ? = \frac{2}{3}$$



Make sure you know how to compute the probabilities above!

Mark as completed



Like



Dislike



Report an issue