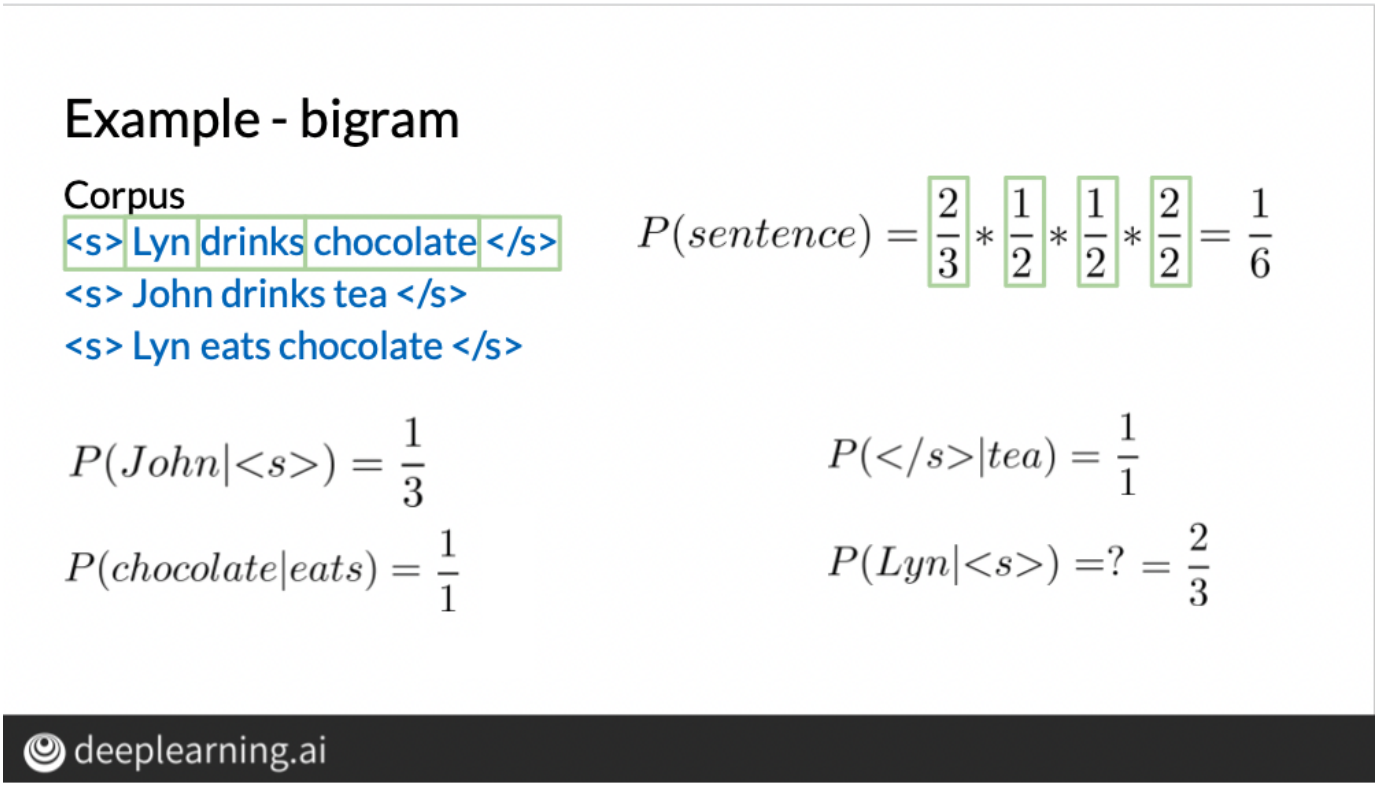
Starting and Ending Sentences

We usually start and end a sentence with the following tokens respectively: <s> </s>.

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

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



Make sure you know how to compute the probabilities above!

Mark as completed