Sequence Probabilities

You just saw how to compute sequence probabilities, their short comings, and finally how to approximate N-gram probabilities. In doing so, you try to approximate the probability of a sentence. For example, what is the probability of the following sentence: *The teacher drinks tea.* To compute it, you will make use of the following:

- $P(B \mid A) = \frac{P(A,B)}{P(A)} \implies P(A, B) = P(A)P(B \mid A)$

- $P(A, B, C, D) = P(A)P(B \mid A)P(C \mid A, B)P(D \mid A, B, C)$

To compute the probability of a sequence, you can compute the following:

$P(\text{ the teacher drinks tea }) = P(\text{the})P(\text{ teacher the})P(\text{ drinks the teacher})P(\text{tea} \mid \text{the teacher drinks })$

One of the main issues with computing the probabilities above is the corpus rarely contains the exact same phrases as the ones you computed your probabilities on. Hence, you can easily end up getting a probability of 0. The *Markov assumption* indicates that only the last word matters. Hence:

- Bigram　　$P(w_n \mid w_1^{n-1}) \approx P(w_n \mid w_{n-1})$

- N-gram　　$P(w_n \mid w_1^{n-1}) \approx P(w_n \mid w_{n-N+1}^{n-1})$

You can model the entire sentence as follows:

- $P(w_1^n) \approx \prod_{i=1}^{n} P(w_i \mid w_{i-1})$

- $P(w_1^n) \approx P(w_1) P(w_2 \mid w_1) \dots P(w_n \mid w_{n-1})$