

Bigram language model

So that's what we get for $n = 2$:

$$p(\mathbf{w}) = \cancel{p(w_1)} p(w_2|w_1) \dots p(w_k|w_{k-1}) \\ p(w_1 | \textit{start})$$