

How to train n-gram models

$$\mathbf{w} = (w_1 w_2 w_3 \dots w_k)$$

Bigram language model:

$$p(\mathbf{w}) = \prod_{i=1}^{k+1} p(w_i | w_{i-1})$$

N-gram language model:

$$p(\mathbf{w}) = \prod_{i=1}^{k+1} p(w_i | w_{i-n+1}^{i-1})$$

$(w_{i-n+1}, \dots, w_{i-1})$

