

Let's do some math

Predict probability of a sequence of words:

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

- **Chain rule:**

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

- **Markov assumption:**

$$p(w_i|w_1 \dots w_{i-1}) = p(w_i|w_{i-n+1} \dots w_{i-1})$$