

Two Views on RNN LM

- RNN is a for loop / functional map over sequential data
- all outputs are conditional distributions
 - probabilistic distribution over sequences of words

$$P(w_1, \dots, w_n) = \prod_{i=1}^n P(w_i | w_{i-1}, \dots, w_1)$$