

Neural Language Models

$$P(\bar{w}) = P(w_1) P(w_2 | w_1) \dots P(w_n | w_1 \dots w_{n-1})$$

model with a neural net

Very basic neural LM: $P(w_i | w_{i-1}) = \frac{e^{v_i \cdot c_{i-1}}}{\sum_{w' \in V} e^{v_w \cdot c_{i-1}}}$

skip-gram

More generally: $P(w_i | w_1 \dots w_{i-1}) = \text{softmax}(U w_i - f(w_1 \dots w_{i-1}))$

f : neural net to embed the context

