Probabilistic Neural Language Model

$$p(w_i | w_{i-n+1}, \ldots w_{i-1}) = \frac{\exp(y_{w_i})}{\sum_{w \in V} \exp(y_w)}$$

Softmax **over components of** *y*

$$y = b + Wx + U \tanh(d + Hx)$$

$$x = [C(w_{i-n+1}), \ldots C(w_{i-1})]^T$$