Probabilistic Neural Language Model

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

Softmax* over components of *y

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

***Feed-forward NN* with tons of parameters**

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

***Distributed representation* of context words**