

Word prediction: skip-gram model

Predict *context words* given a focus word:

$$p(w_{i-h}, \dots, w_{i+h} | w_i) = \prod_{-h \leq k \leq h, k \neq 0} p(w_{i+k} | w_i)$$

Model each probability with a *softmax*:

$$p(u|v) = \frac{\exp \langle \phi_u, \theta_v \rangle}{\sum_{u' \in W} \exp \langle \phi_{u'}, \theta_v \rangle}$$

Still two matrices of parameters.