Skip-gram Negative Sampling (SGNS)

Instead of predicting a word for another word, predict "yes" or "no" for word pairs:

$$\sum_{u \in W} \sum_{v \in W} n_{uv} \log \sigma \left(\langle \phi_u, \theta_v \rangle \right) +$$

$$k \, \mathbb{E}_{\bar{v}} \log \sigma \left(- \langle \phi_u, \theta_{\bar{v}} \rangle \right) \to \max_{\phi_u, \theta_v}$$

- Use positive examples from data: v co-occurred with u
- Sample negative examples: k random $\bar{v}$ from the vocabulary

Train with SGD to find two matrices of parameters (as usual).