Skip-Gram Negative Sampling Objective

Locally,

$$\ell = \log \sigma(\mathbf{w} \cdot \tilde{\mathbf{c}}) + k \cdot \mathbb{E}_{c_N \sim p_D(c_N)} \left[\log \sigma(-\mathbf{w} \cdot \tilde{\mathbf{c_N}})\right]$$

and thus globally

$$\mathcal{L} = \sum_{w \in V} \sum_{c \in V} (n_{w,c}) \left(\log \sigma(\mathbf{w} \cdot \tilde{\mathbf{c}}) + k \cdot \mathbb{E}_{c_N \sim P_n(c_N)} \left[\log \sigma(-\mathbf{w} \cdot \tilde{\mathbf{c_N}})\right]\right)$$

- k: number of negative samples to take (hyperparameter)
- $n_{w,c}$: number of times (w, c) was seen in the data
- $\mathbb{E}_{c_N \sim P_n(c_N)}$ indicates an expectation taken with respect to the noise distribution $P_n(c_N)$.