

Relation to word2vec

GloVe Objective:

$$J = \sum_{i,j=1}^V f(X_{ij})(\mathbf{w}_i^T \tilde{\mathbf{w}}_j + b_i + \tilde{b}_j - \log X_{ij})^2$$

word2vec (Skip-Gram) Objective (after rewriting):

$$J = - \sum_{i=1}^V X_i \sum_{j=1}^V P(w_j | w_i) \log Q(w_j | w_i)$$

where $X_i = \sum_k X_{ik}$ and P and Q are the empirical co-occurrence and model co-occurrence distributions, respectively. **Weighted cross-entropy error!**

Authors show that **replacing cross-entropy with least-squares nearly re-derives GloVe:**

$$\hat{J} = \sum_{i,j} X_{ij} (\mathbf{w}_i^T \tilde{\mathbf{w}}_j - \log X_{ij})^2$$