

Word2Vec: Two Training Objectives

Skip-gram Objective

Given: Center word

Predict: Context words

$$\max \sum_t \sum_{j \neq 0} \log P(w_{t+j} | w_t)$$

Better for: Rare words

More common in practice

CBOW Objective

Given: Context words

Predict: Center word

$$\max \sum_t \log P(w_t | context_t)$$

Better for: Frequent words

Faster training