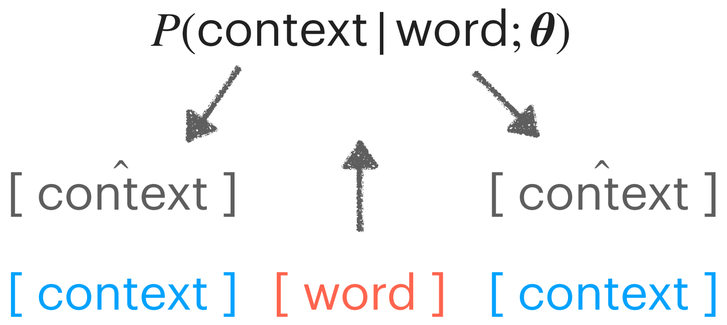
The skip gram model



Example: the quick brown fox jumped over the lazy dog

$\mathbf{x}_{t-c} \quad \cdots \quad \mathbf{x}_t \quad \cdots \quad \mathbf{x}_{t+c}$

Learning problem:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \prod_{t=1}^{T} \prod_{-c \le j \le c, j \neq 0} \mathbf{x}_{t+j} \cdot p(\mathbf{x}_{t+j} \mid \mathbf{x}_t; \boldsymbol{\theta})$$

$$= \arg\min_{\boldsymbol{\theta}} - \sum_{t=1}^{T} \sum_{-c \le j \le c, j \neq 0} \mathbf{x}_{t+j} \cdot \log p(\mathbf{x}_{t+j} \mid \mathbf{x}_t; \boldsymbol{\theta})$$

where

T = corpus length

c = context window

N = vocabulary size

$\mathbf{x} \in \{0,1\}^N$

$\boldsymbol{\theta}$ = parameterization

Word2Vec: learn a feature representation that preserves semantic relationships between words based on distance

Produced using Word2Vec word vector representations, compressed to 2D using tSNE