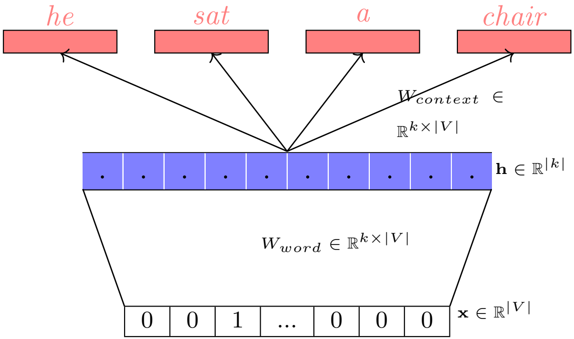
Module 10.10: Relation between SVD & word2Vec

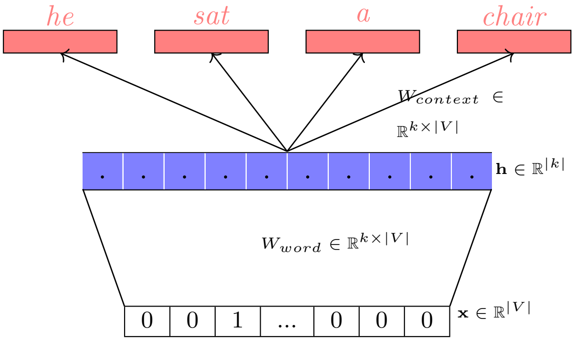
The story ahead ...

- Continuous bag of words model
- Skip gram model with negative sampling (the famous word2vec)
- GloVe word embeddings
- Evaluating word embeddings
- Good old SVD does just fine!!



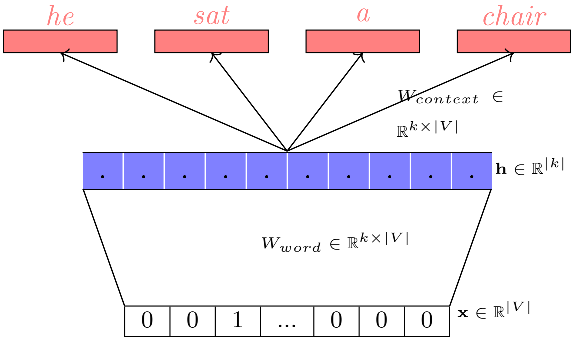
- Recall that SVD does a matrix factorization of the co-occurrence matrix

.



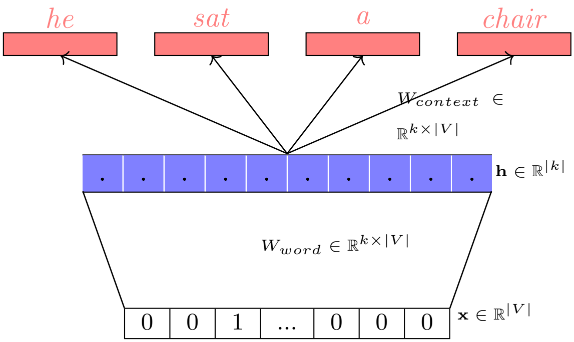
- Recall that SVD does a matrix factorization of the co-occurrence matrix
- Levy et.al [2015] show that word2vec also implicitly does a matrix factorization

.

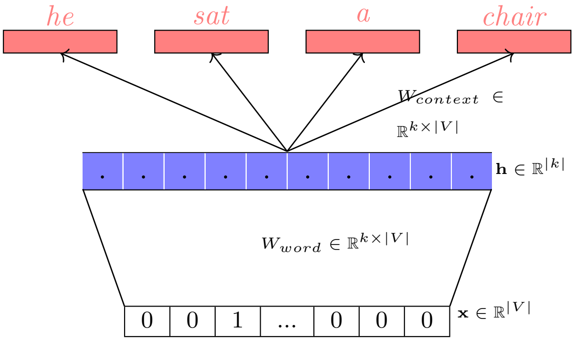


- Recall that SVD does a matrix factorization of the co-occurrence matrix
- Levy et.al [2015] show that word2vec also implicitly does a matrix factorization
- What does this mean ?

.



- Recall that SVD does a matrix factorization of the co-occurrence matrix
- Levy et.al [2015] show that word2vec also implicitly does a matrix factorization
- What does this mean ?
- Recall that word2vec gives us $W_{context}$ & W_{word} .



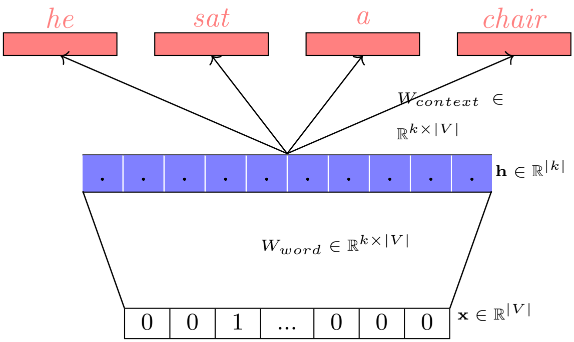
- Recall that SVD does a matrix factorization of the co-occurrence matrix
- Levy et.al [2015] show that word2vec also implicitly does a matrix factorization
- What does this mean ?
- Recall that word2vec gives us $W_{context}$ & W_{word} .
- Turns out that we can also show that

$$M = W_{context} * W_{word}$$

where

$$M_{ij} = PMI(w_i, c_i) - log(k)$$

$$k = \text{number of negative samples}$$



- Recall that SVD does a matrix factorization of the co-occurrence matrix
- Levy et.al [2015] show that word2vec also implicitly does a matrix factorization
- What does this mean ?
- Recall that word2vec gives us $W_{context}$ & W_{word} .
- Turns out that we can also show that

$$M = W_{context} * W_{word}$$

where

$$M_{ij} = PMI(w_i, c_i) - log(k)$$
$$k = \text{number of negative samples}$$

- So essentially, word2vec factorizes a matrix M which is related to the PMI based co-occurrence matrix (very similar to what SVD does)