Word2vec (1)

Pretrained NLP network as startpoint for specialization.
Has learned concepts (comparable to CNN for images)

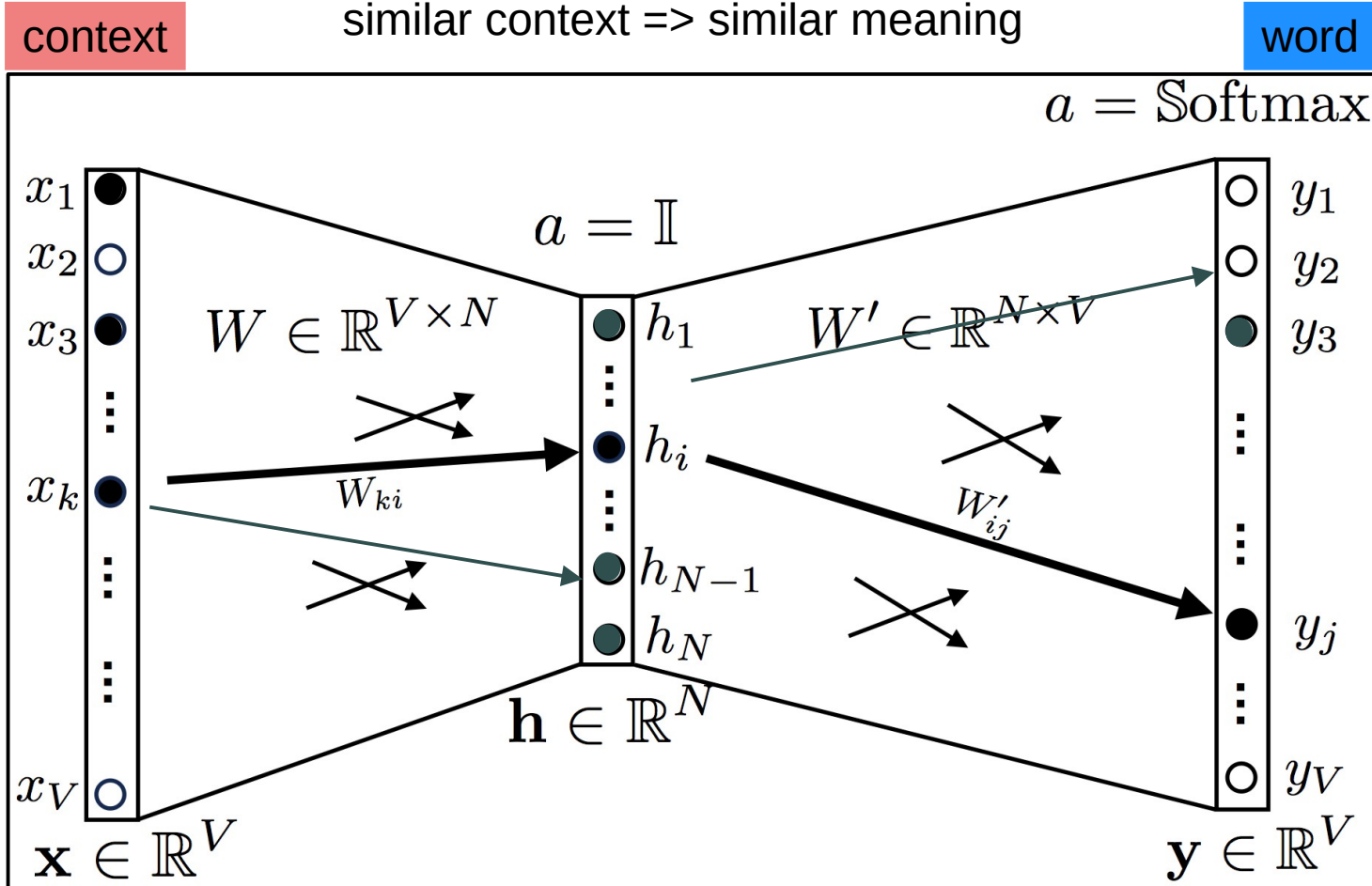[blue box] : Center Word

[red box] : Context Word

c=0   The cute cat jumps over the lazy dog.

c=1   The cute cat jumps over the lazy dog.

c=2   The cute cat jumps over the lazy dog.

word2vec (2)

similar context => similar meaning

context   **word**

$a = \mathbb{Softmax}$

x_1 y_1

$a = \mathbb{I}$

x_2 y_2

x_3 $W \in \mathbb{R}^{V \times N}$ h_1 $W' \in \mathbb{R}^{N \times V}$ y_3

x_k W_{ki} h_i W'_{ij}

h_{N-1}

h_N y_j

$\mathbf{h} \in \mathbb{R}^N$

x_V y_V

$\mathbf{x} \in \mathbb{R}^V$ $\mathbf{y} \in \mathbb{R}^V$
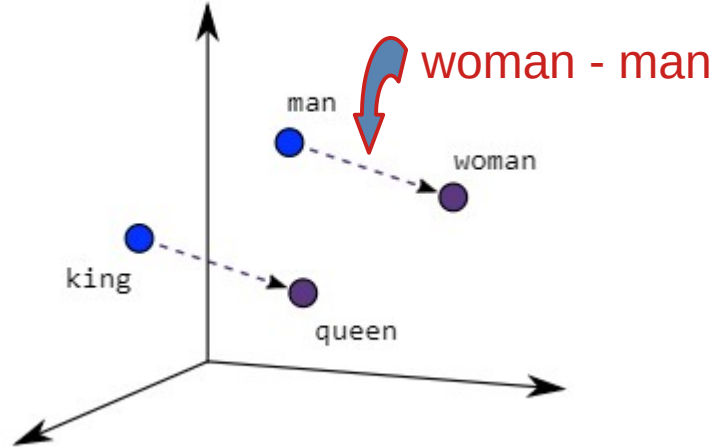
Dimension reduction:

Similar contexts should produce synonyms, hence be reduced to similar vectors by efficient coding in the hidden layer.

CBOW (Continuous Bag Of Words)
Inverse: Skip-gram

Zellig S Harris:
Distributional Structure (1956)

word2vec (3)

king – man + woman = queen

Bert (Google)

Trouble with word2vec:

house (to live in)
house (music)

Solution: Bert takes bi-directional context into account.

Bidirectional
Encoder
Representations from
Transformers