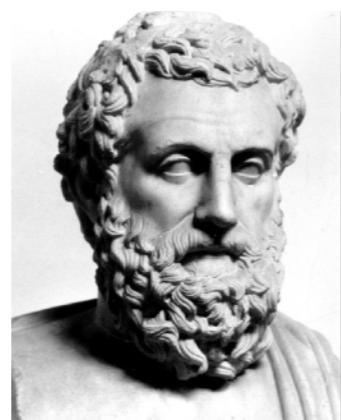


# Artificial Intelligence For NLP Lesson- 11

人工智能与自然语言处理  
课程组

2019.Sept. 14

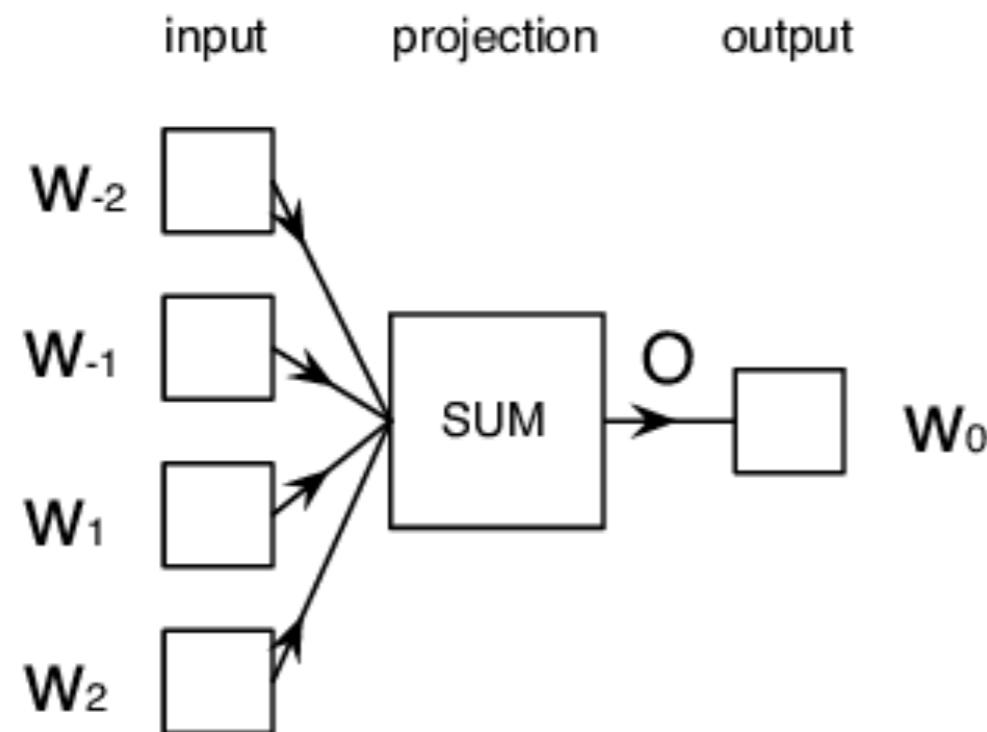


## Outline

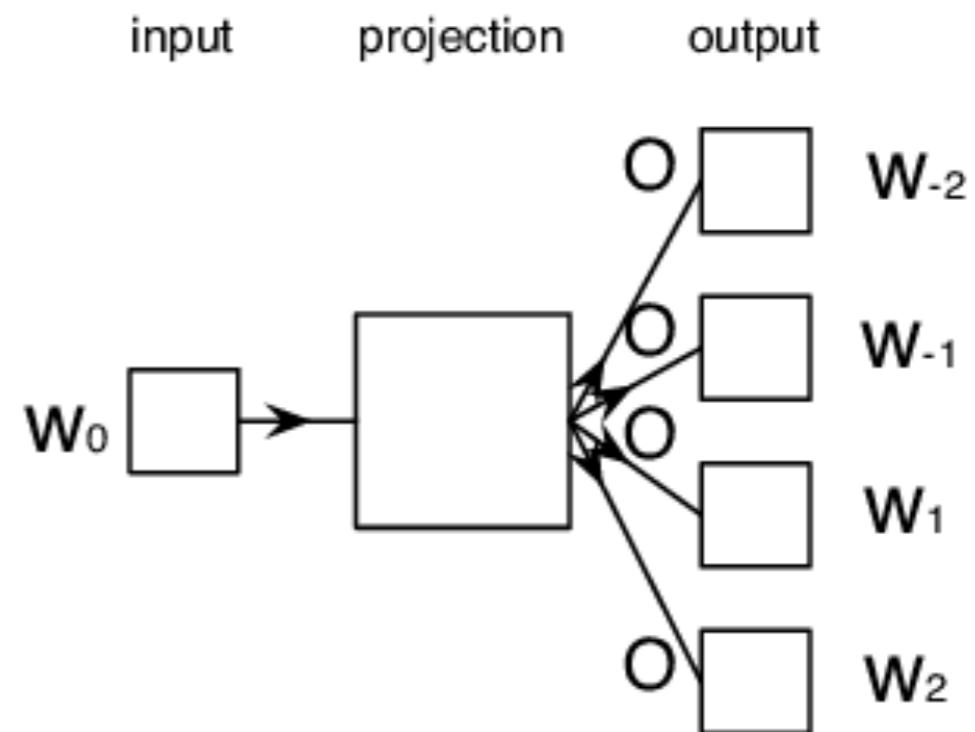
- 1. Review
  - What's the graph computing?
  - What's the neural networks framework?
- 2. Word2Vec Advanced Topic
  - Hierarchy Softmax
  - Negative Sampling
  - Glove
  - Fasttext

## CBOW

### CBOW

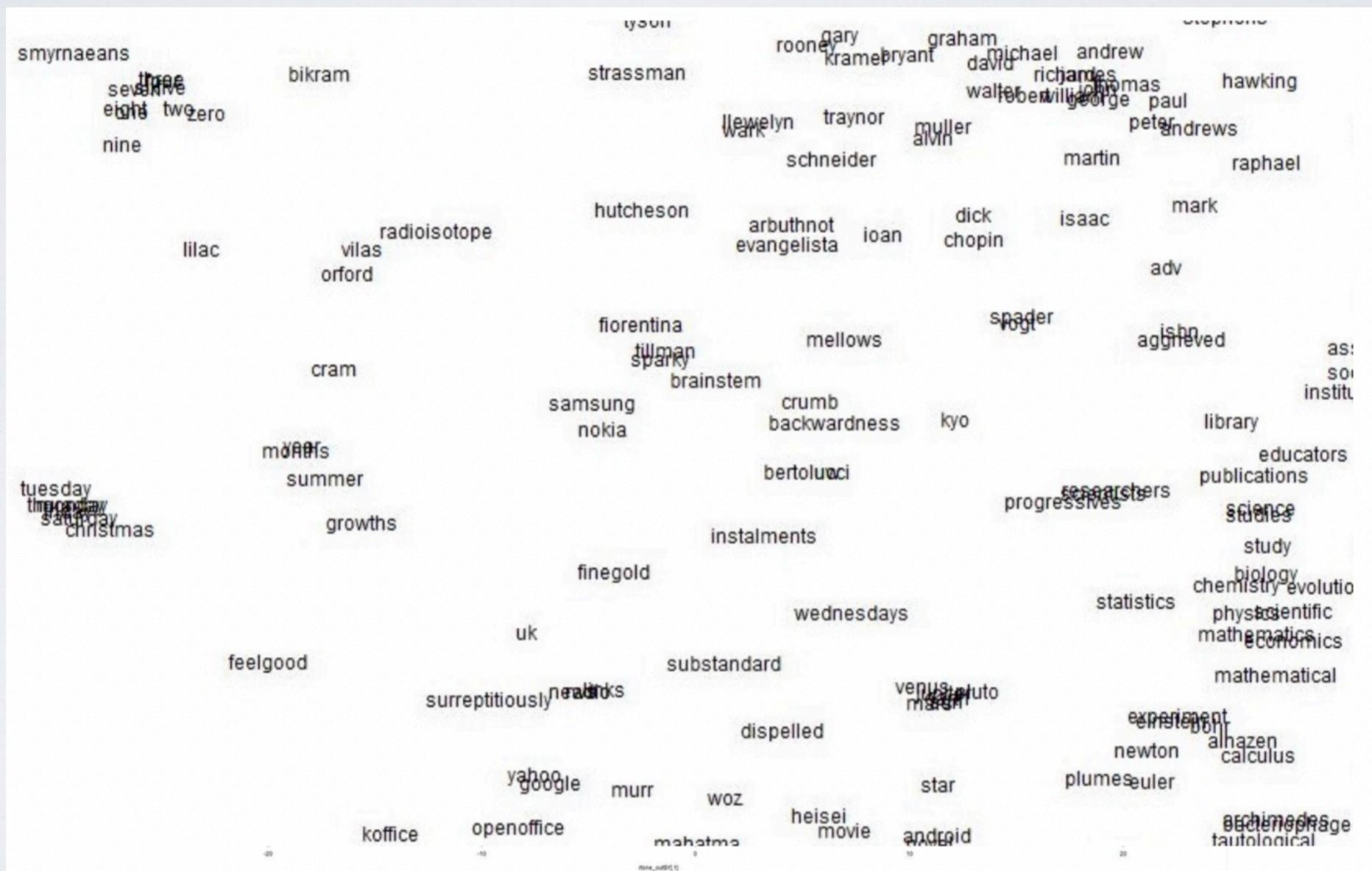


### Skip-Ngram



$$p(w_j|w_I) = y_j = \frac{\exp(u_j)}{\sum_{j'=1}^V \exp(u_{j'})},$$

$$J(\theta) = \frac{1}{T} \sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log p(w_{t+j}|w_t)$$



In [7]:

```
model.wv.most_similar('勇敢')
```

```
[('勇于', 0.5452967882156372),
 ('坚毅', 0.544731855392456),
 ('坚强', 0.5447058081626892),
 ('勇气', 0.537638783454895),
 ('果敢', 0.5369020700454712),
 ('善良', 0.5327401757240295),
 ('坚忍不拔', 0.5091272592544556),
 ('豁达', 0.5033485293388367),
 ('真诚', 0.5024720430374146),
 ('追爱', 0.4999522566795349)]
```

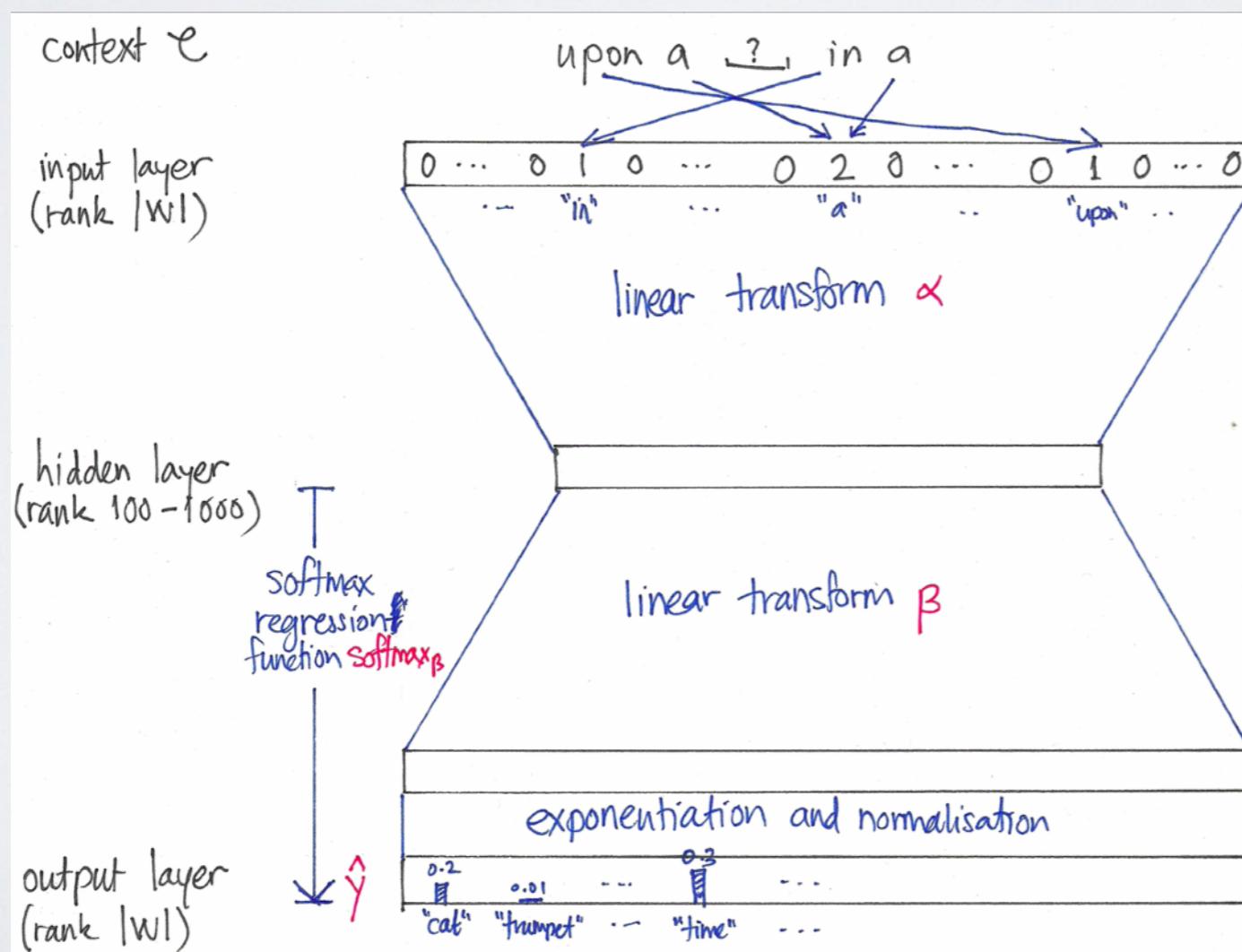
In [31]:

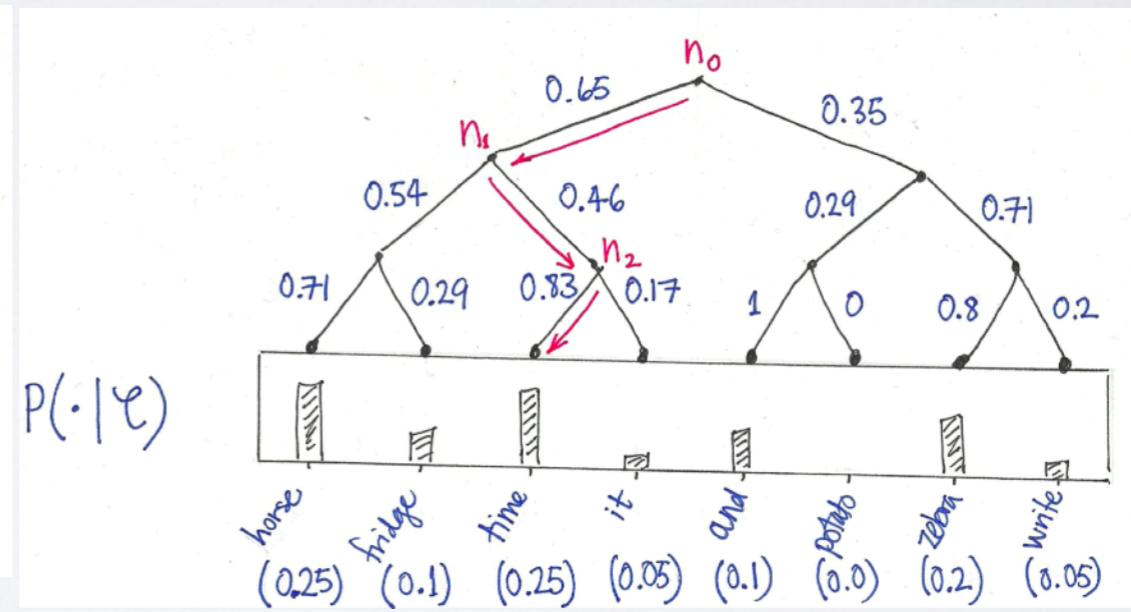
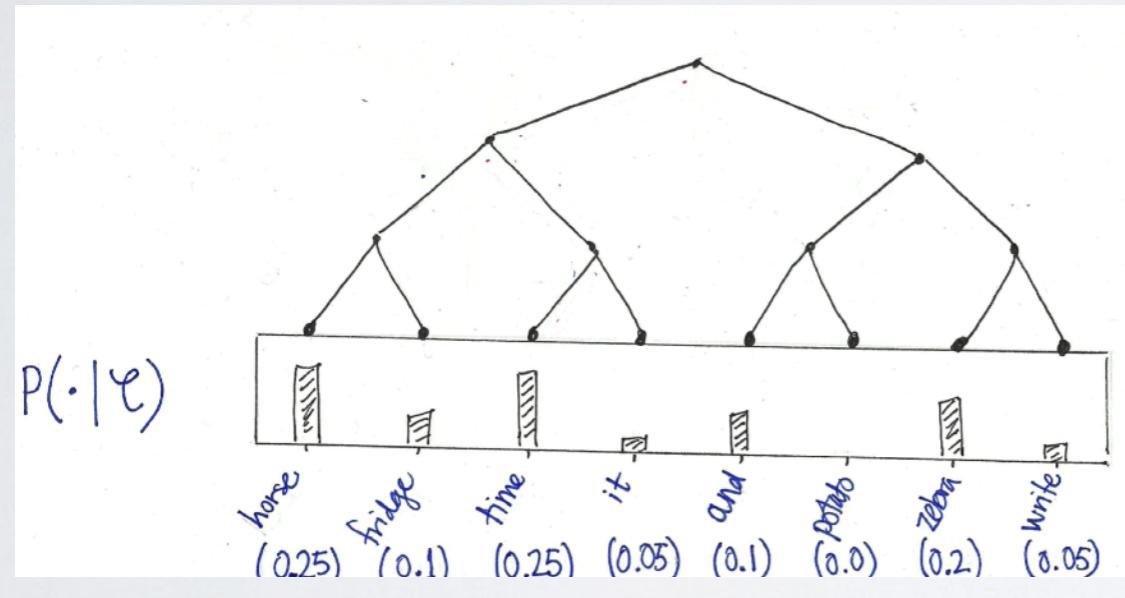
```
model.wv.most_similar('美女')
```

```
[('帅哥', 0.5578837990760803),
 ('校花', 0.5232111215591431),
 ('女演员', 0.5189783573150635),
 ('金发碧眼', 0.5069225430488586),
 ('超模', 0.502299964427948),
 ('舞女', 0.5002666711807251),
 ('甜心', 0.4993056654930115),
 ('男模', 0.4970622658729553),
 ('女主播', 0.49386513233184814),
 ('嫩模', 0.4927775263786316)]
```

## hierarchical softmax

- Vocabulary is so huge. Softmax need a unacceptable time.





## Huffman Tree

word	count
fat	3
fridge	2
zebra	1
potato	3
and	14
in	7
today	4
kangaroo	2

→

Huffman tree

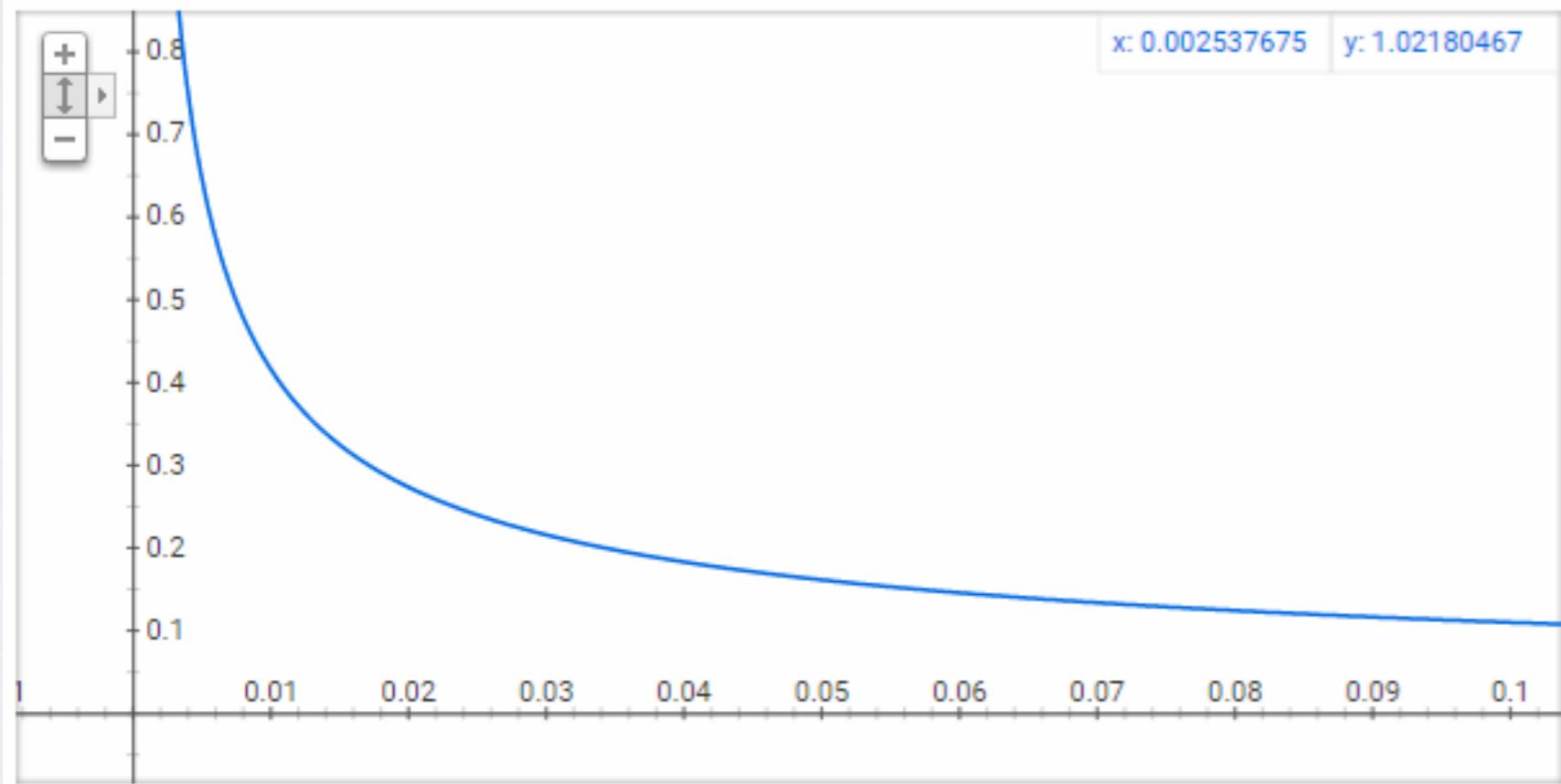
```
graph TD; and[and] --> today[today]; and --> in[in]; today --> fridge[fridge]; today --> kangaroo[kangaroo]; in --> fat[fat]; in --> potato[potato]
```

## Negative sampling

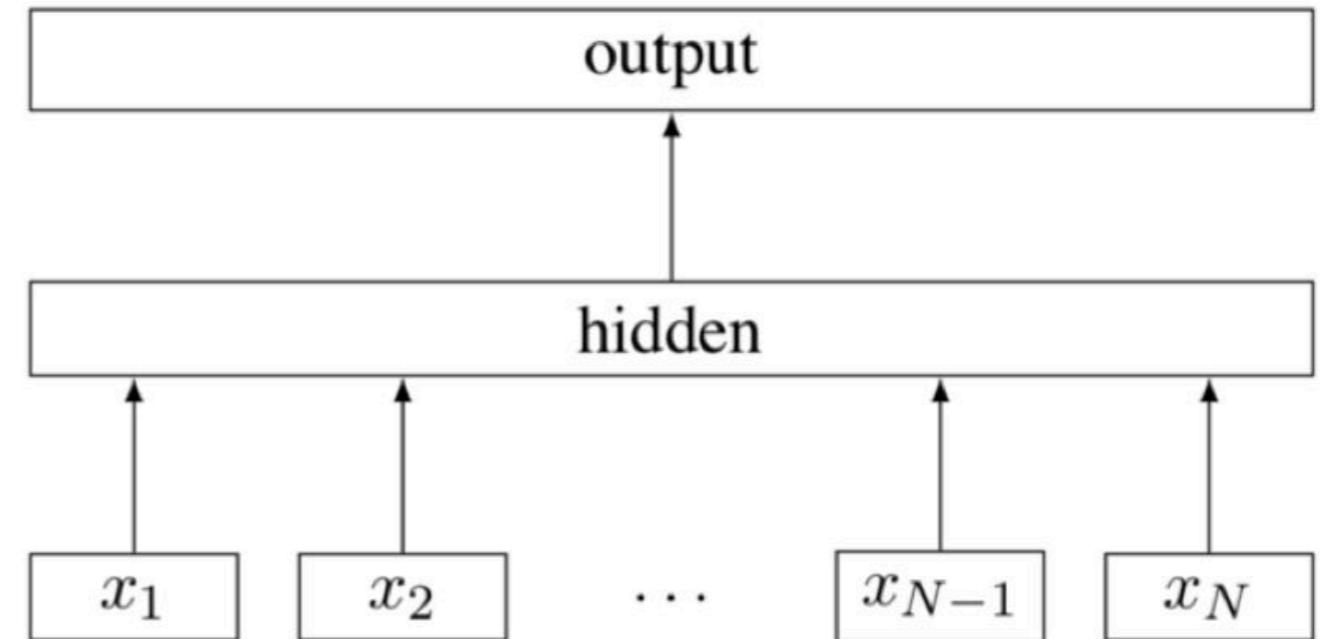
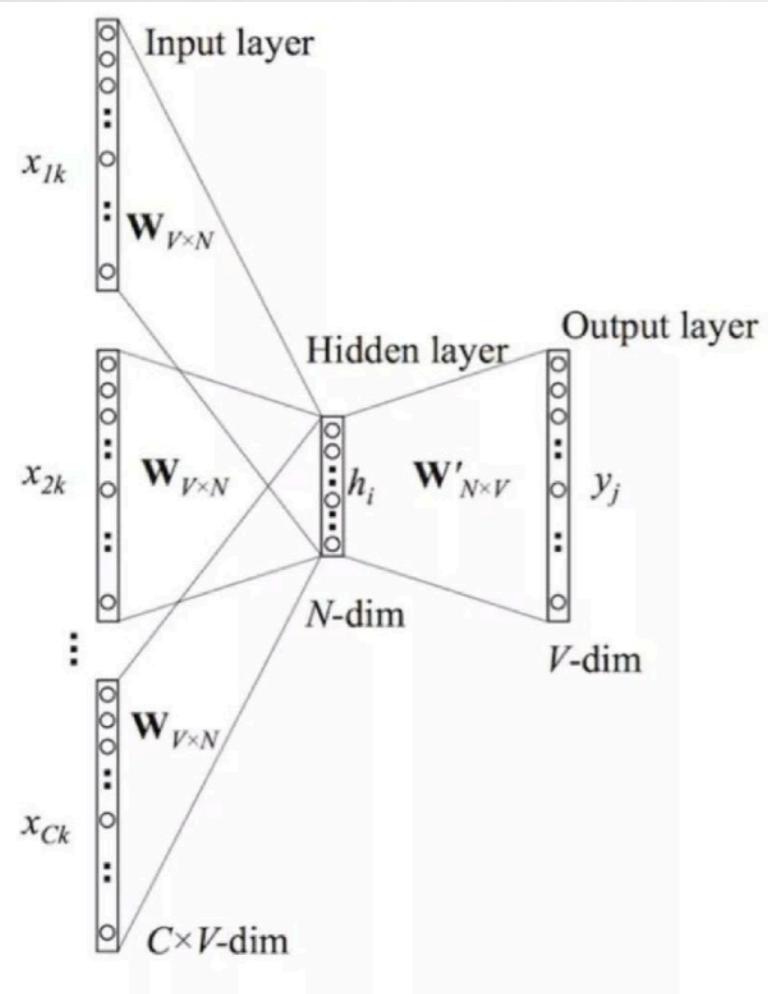
$$J_t(\theta) = \log \sigma(u_o^T v_c) + \sum_{j \sim P(w)} [\log \sigma(-u_j^T v_c)]$$

Graph for  $(\sqrt{x/0.001}+1)$ :

$$P(w_i) = (\sqrt{\frac{z(w_i)}{0.001}} + 1) \cdot \frac{0.001}{z(w_i)}$$



## FastText



**Figure 1:** Model architecture of `fastText` for a sentence with  $N$  ngram features  $x_1, \dots, x_N$ . The features are embedded and averaged to form the hidden variable.

- e.g: 阿里巴巴和蚂蚁金服集团

# Glove

	the	cat	sat	on	mat
the	0	1	0	1	1
cat	1	0	1	0	0
sat	0	1	0	1	0
on	1	0	1	0	0
mat	1	0	0	0	0

Probability and Ratio	$k = \text{solid}$	$k = \text{gas}$	$k = \text{water}$	$k = \text{fashion}$
$P(k \text{ice})$	$1.9 \times 10^{-4}$	$6.6 \times 10^{-5}$	$3.0 \times 10^{-3}$	$1.7 \times 10^{-5}$
$P(k \text{steam})$	$2.2 \times 10^{-5}$	$7.8 \times 10^{-4}$	$2.2 \times 10^{-3}$	$1.8 \times 10^{-5}$
$P(k \text{ice})/P(k \text{steam})$	8.9	$8.5 \times 10^{-2}$	1.36	0.96

## Assignment

- 1. Efficient Estimation of Word Representations in Vector Space, <https://arxiv.org/pdf/1301.3781.pdf>
- 2. Reading:Distributed Representations of Words and Phrases and their Compositionality(<http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf>)