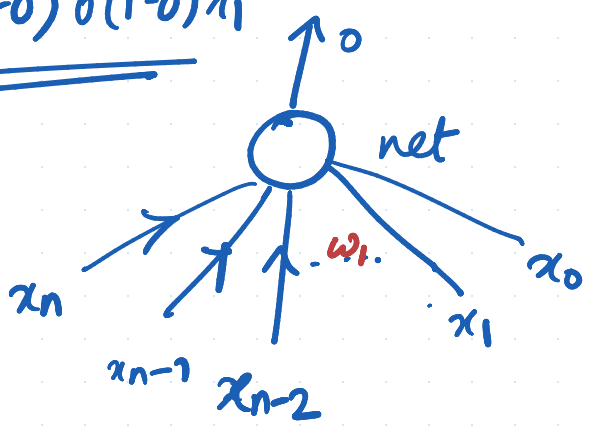
Lecture - 5 (Theory + Derivations)

Single Neuron : Sigmoid + total sum of square (TSS) loss.

$$\Delta w_1 = -\eta \frac{\partial L}{\partial w_1}$$

η = Learning rate.

$$= +\eta\,(t-o)\,o(1-o)\,x_1$$



$$\left(L = \text{loss} = \frac{1}{2}(t-o)^2 \right)$$

t = target , o = output.

$$\frac{\partial L}{\partial o} = \frac{1}{2} \times 2 \times (t-o)(-1)$$

$$\boxed{\frac{\partial L}{\partial o} = -(t-o)}$$

$$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial o} \cdot \frac{\partial o}{\partial net} \cdot \frac{\partial net}{\partial w_1}$$

$$net = \sum_{i=0}^{n} w_i x_i$$

$$\boxed{o = \frac{1}{1+e^{-net}}}$$

$$\frac{\partial o}{\partial net} = o(1-o)$$

$$\frac{\partial net}{\partial w_1} = x_1$$

$$= -(t-o) \cdot o(1-o) \cdot x_1$$

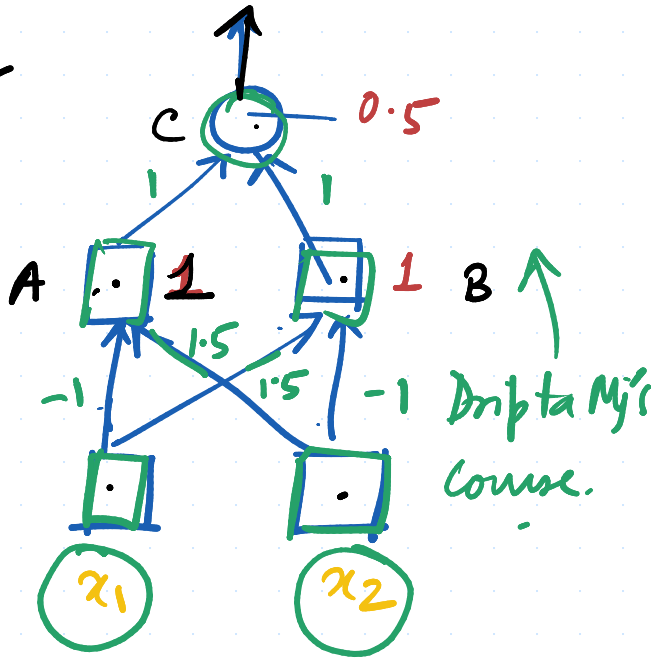
XOR → classification network.

$x_1\bar{x}_2 + x_2\bar{x}_1$

x_1	x_2	$x_1 \oplus x_2$
0	0	0
0	1	1
1	0	1
1	1	0

Neurons ⇒ 5

Connections ⇒ 6



Drifta Mj's Course.

A

A's o/p

0 0 → 0 ≥ 1 ? → 0

0 1 → 1·5 ≥ 1 ? → 1

1 0 → −1 ≥ 1 ? → 0

1 1 → 0·5 ≥ 1 ? → 0.

x_1 x_2

B

0 0 → 0 ≥ 1 → 0

0 1 → −1 ≥ 1 → 0

1 0 → 1·5 ≥ 1 → 1

1 1 → 0·5 ≥ 1 → 0.

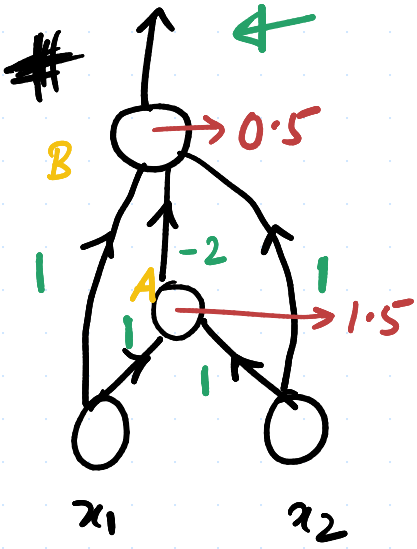
x_1 x_2

A	B	
0	0	0 ≥ 0·5 → 0
1	0	1 ≥ 0·5 → 1
0	1	1 ≥ 0·5 → 1
0	0	0 ≥ 0·5 → 0

x_1	x_2	$x_1 \oplus x_2$
0	0	0
0	1	1
1	0	1
1	1	0

Neurons = 4

connections ⇒ 5



x_1	x_2	A's output
0	0	0
0	1	0
1	0	0
1	1	1

→

$1 \times 1 + 1 \times 1 = 2 > 1.5$?

⇒ 1

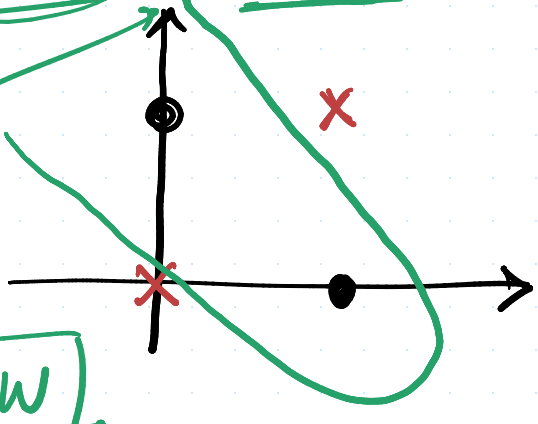
→

x_1	x_2	A o/p	B's
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

find out.

(?) (?), Weights ?

- Quadratic output → Arch, Neurons
- Single sigmoid neuron with sin-function $\left(\frac{\pi}{2}, \pi-\theta\right)$ Can compute (XOR)



H/W

- Multiple Neurons formig more than one layer with linear o/p → collapses to a single neuron

Net linear ← $(k_1 x_1 + k_2 x_2 + k_3)$
comb.

NLP.

Words forms meaning by its association with other words.

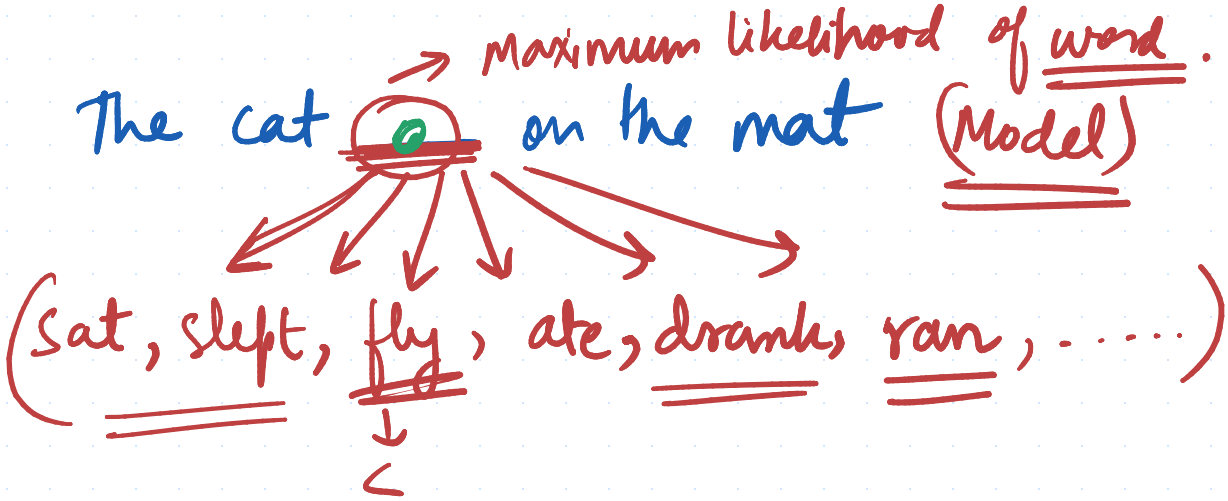
Harris Distributional Property | Lexicographical way → sort.

NLP → words.

(Corpus)
text → dump.

One-hot-encoding

wikepedia page → words, Sentences, paragraph.



Auto-complete

The sky is (blue).

data → (30GB)

similar words → close together.

word → 'characters' →

↓

numerical system → Distance computation b/w words.

↓

$\begin{bmatrix} 1.5 \\ 3.6 \\ -7.6 \end{bmatrix}$

vocabulary → (2.5 lacs)

2.5 lacs

$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ← ate.

bite $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$\begin{bmatrix} \text{ate} \\ \text{bite} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$

(waste of space)

$< \cdot >$ cosine similarity distance.

$\Rightarrow 0$. ✗ not true

(ate, eat) $\Rightarrow 0$ → similar words

(2.5×2.5) lacs

[0 ⋮ 0] → [ ] 1-10%

2.5 lac. → (2048)

25000

Chat GPT

↓

(Bharat ek prodhan matri ke?)

Bharat ek ph

Code Mix → Algo?

5 years → experience

2020 CV

Song generate

Bi-direc LSTM.

NLP ×

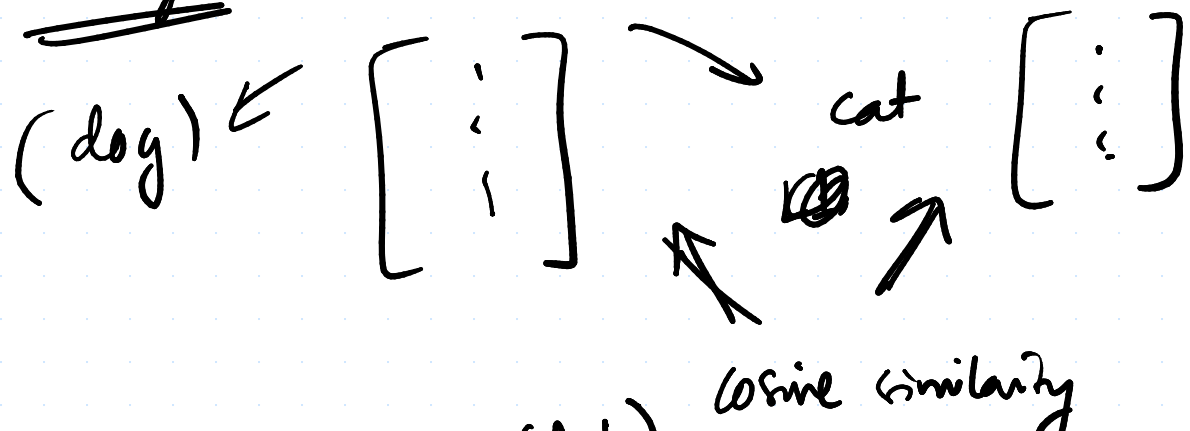
2-3

'bengali' → (phonemes) → 'english' → (speech)

sound

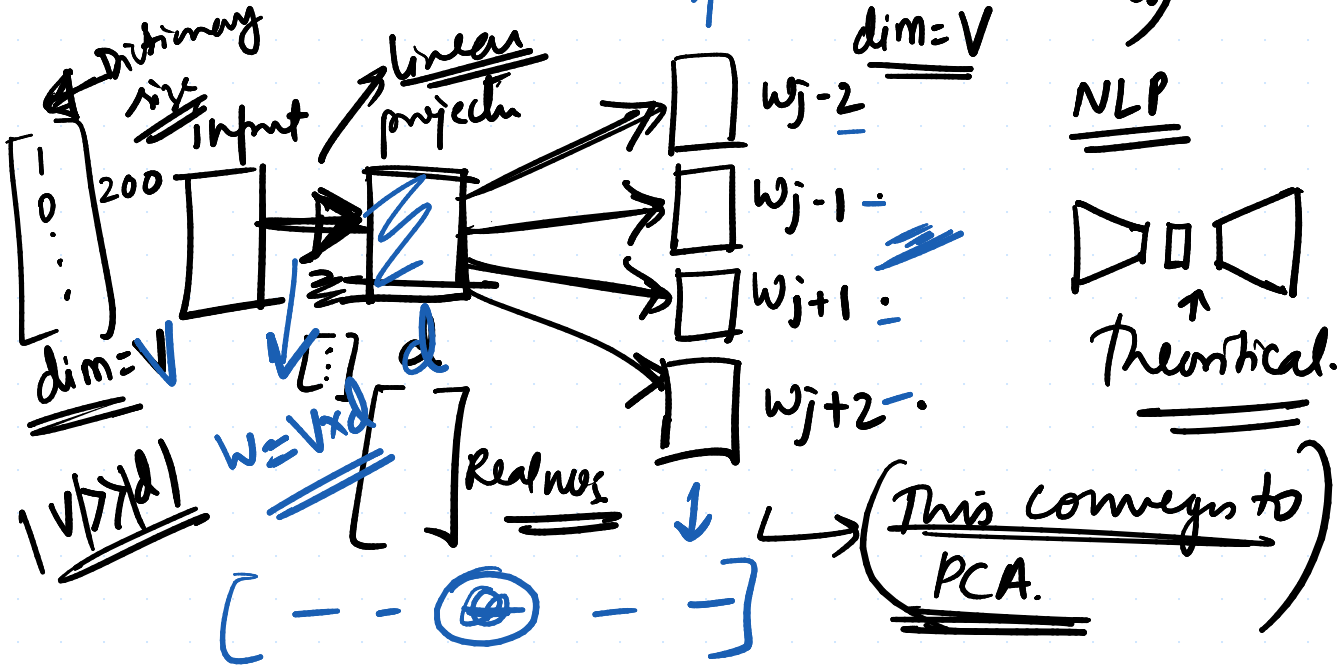
MFCC encoding

- Similar words should have similar embeddings.



Word-2-Vec (Thomas-Mikolov)

Skip-gram



CBOW (Continuous Bag of words)

NLP

Theoretical.

This converges to PCA.

Learning Algorithm → (unsupervised)

Corpus

{The, sky, is, blue}
sky

{the, sky, is, blue}

{the, sky, is, blue, the}

lexicographical

{blue, is, sky, the}

Corpus
unique

[: : : : : :]

SVD →

$M = U \Sigma V^*$

$m \times n$ = $m \times m$ $m \times n$ $n \times n$

Eigenvalues → Diagonal

PCA

$m \times m$
~~$m \times n$~~
~~$n \times n$~~

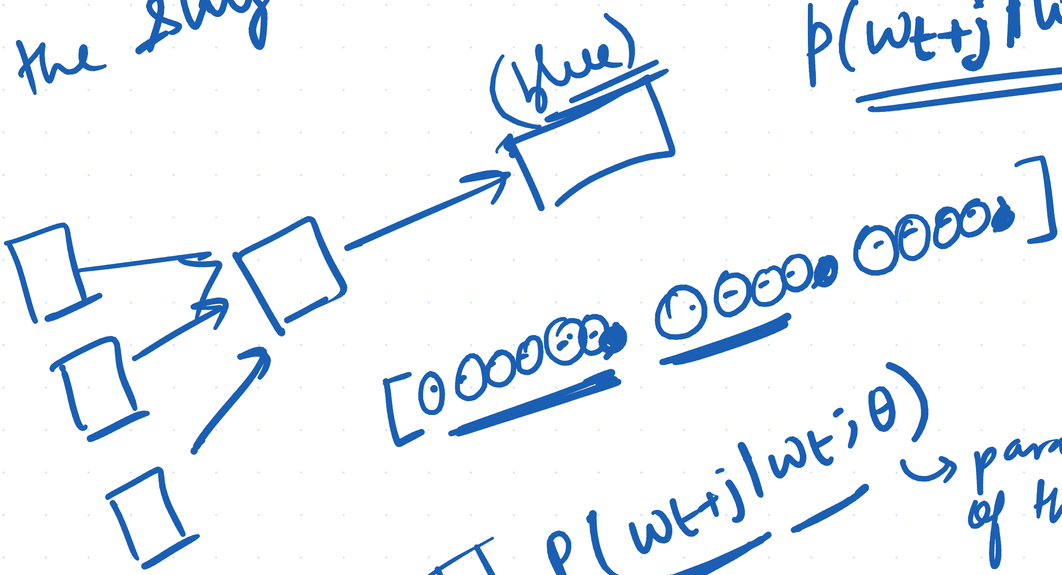
(cd) $(c = \frac{1}{4} d)$

CBOW

The sky is ___ (blue)

Modelling.

$p(w_{t+j} | w_t)$

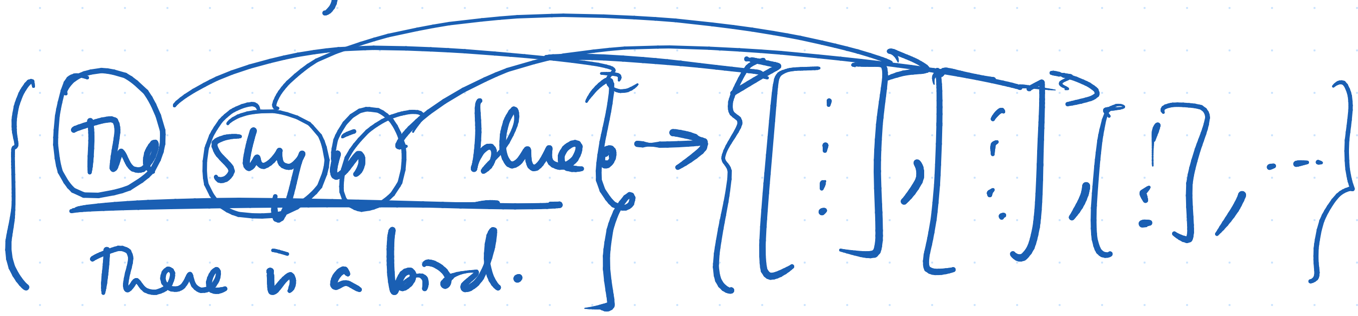


$$J'(\theta) = \frac{1}{T} \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \ne 0}} p(w_{t+j} | w_t ; \theta)$$

→ parameter of the model.

MLE

$$J(\theta) = -\frac{1}{T} \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \ne 0}} p(w_{t+j} | w_t ; \theta)$$



- - - - - []

[The wide road shimmered] in the hot sun.

skip gram. [wide, the
wide, road
wide, shimmered]

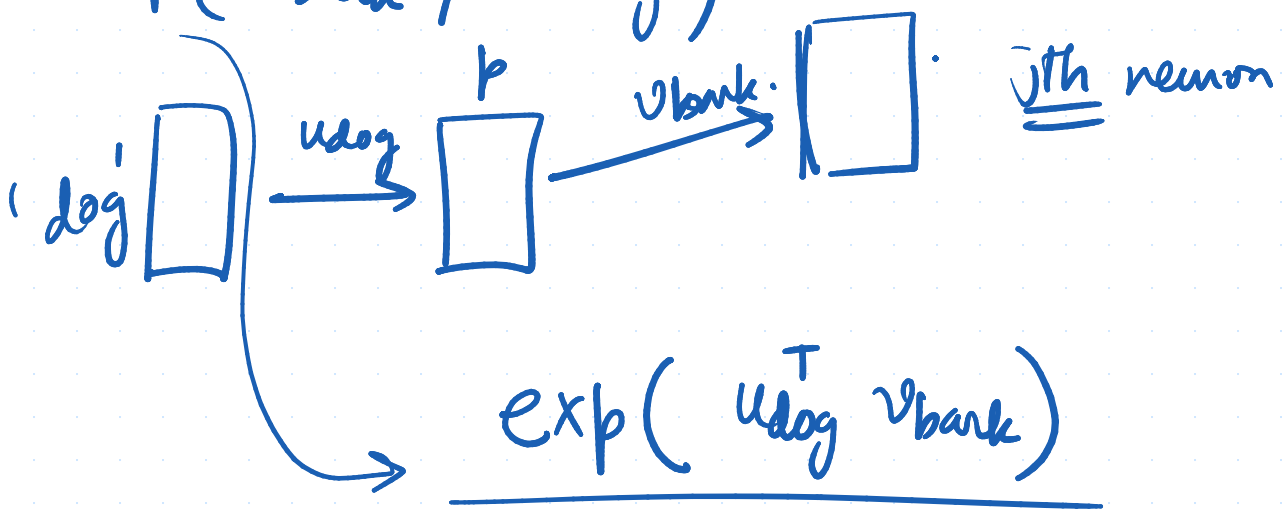
The [wide road shimmered in the] hot sun

[shimmered, wide
shimmered, road
shimmered, in
shimmered the]

minimize $\mathcal{L} = -\sum_{t=1}^{T} \sum_{-m \le j \le m,\ j \neq 0} \log [p(w_{t+j} / w_{tj})]$

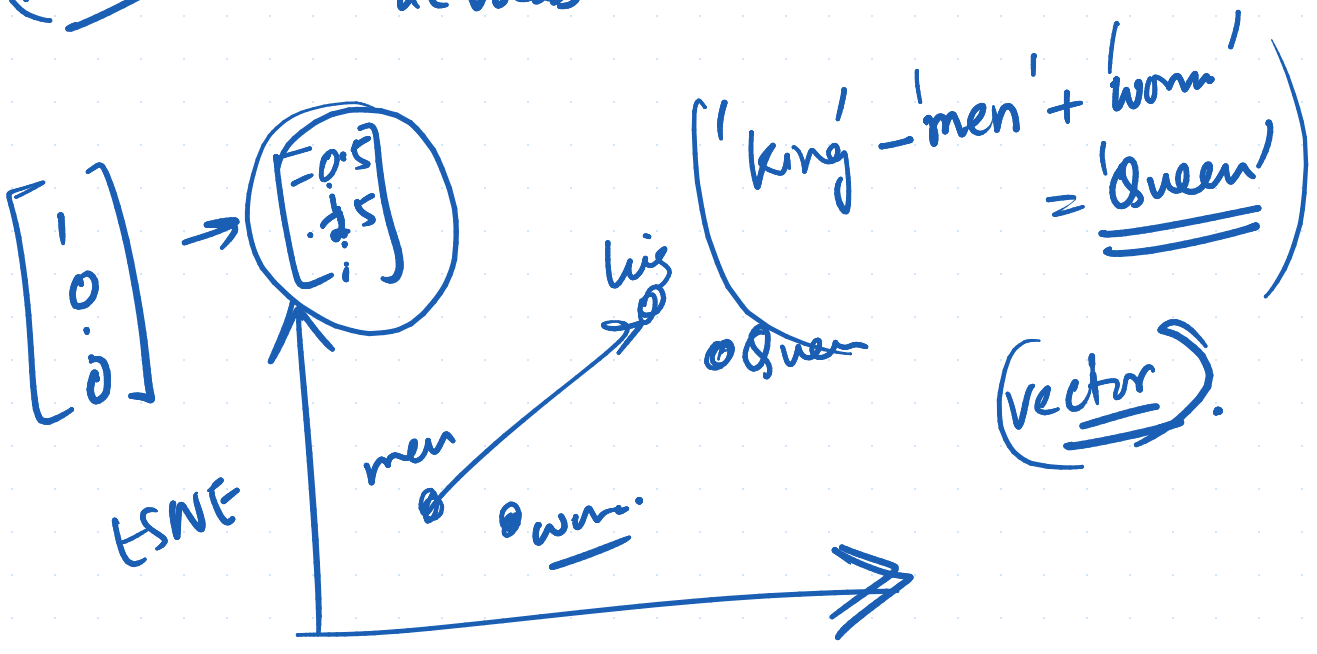
P('context word' | 'input word')

P('bank' | 'dog')



$$\frac{\exp(u_{dog}^T v_{bank})}{\sum_{v_k \in vocab} \exp(u_{dog}^T v_k)}$$

Probability (Model)



Word → numerical value.