

Today's

- 1 Vector -
- 2 Vector Similarity

↳ Cosine Similarity

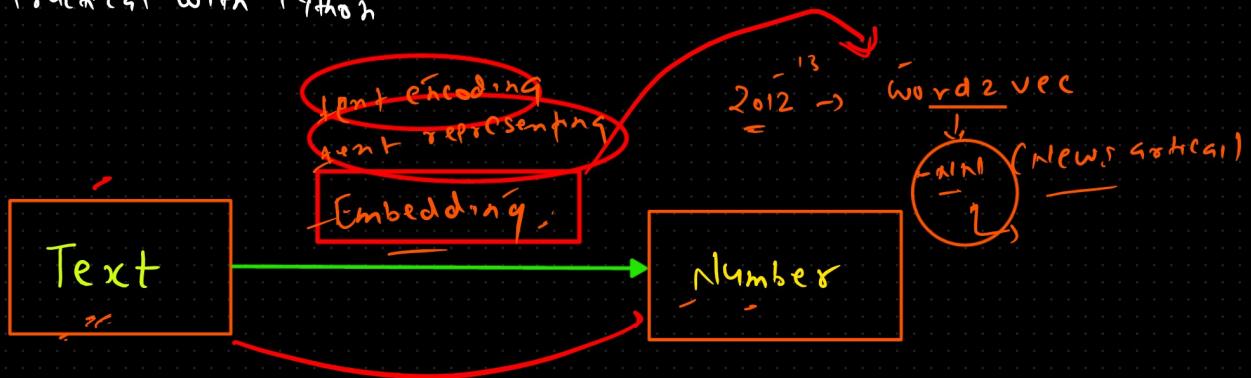
↳ Dot Product

- 3 Embedding → word2vec

↳ CBOW

↳ Skip gram

- 4 Practical with Python



type

frequency

1. BOW

2. n-gram

3. TF-IDF

4. glove (global vectors)
(Matrix factorization)

Prediction

{ word2vec
→ CBOW
→ skipgram }

Neural network

tomorrow

→ Attention all you need

Word2Vec → Word to vector

1 Semantic meaning

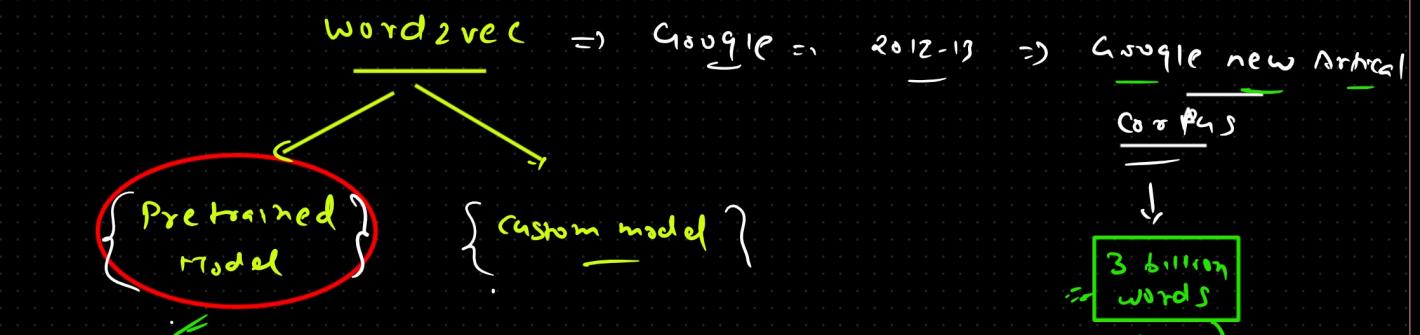
TF-IDF, BOW → N.99m

happy, joy

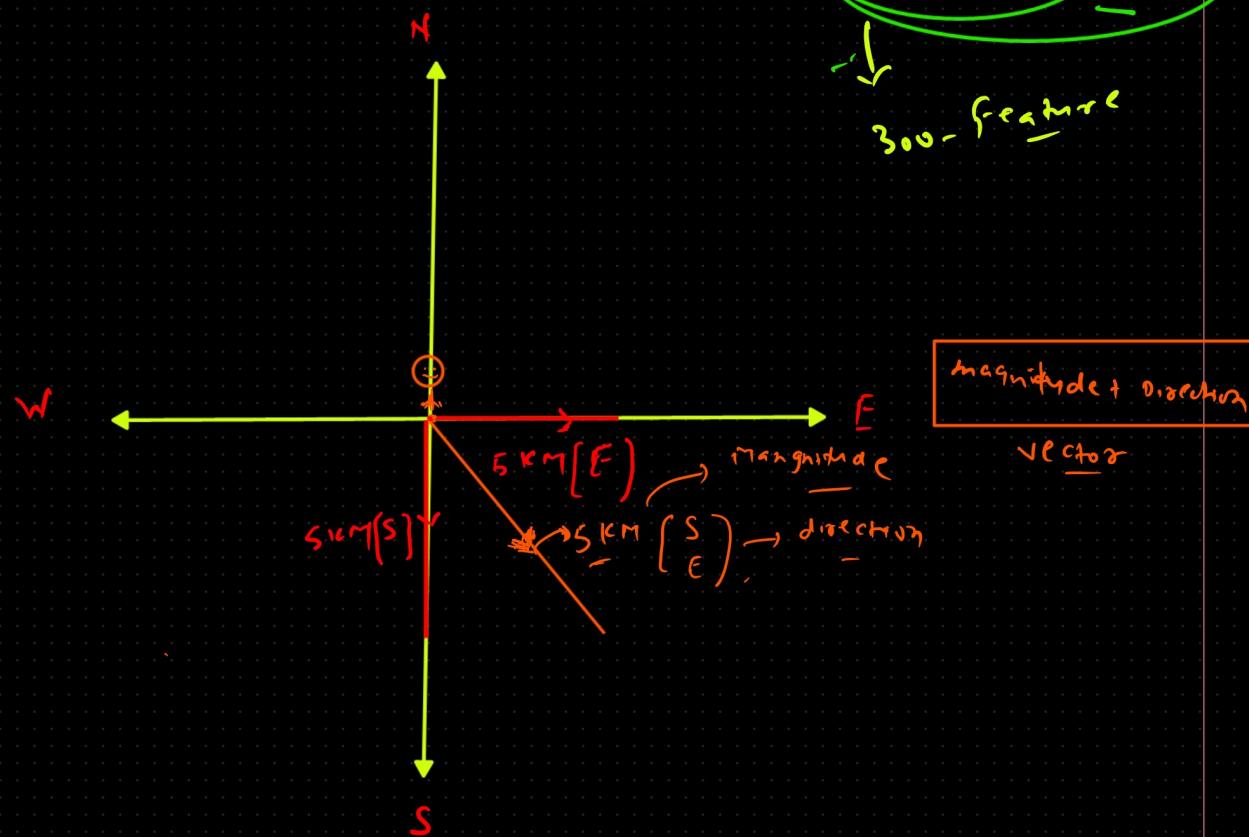
2 higher dim. → lower dim.

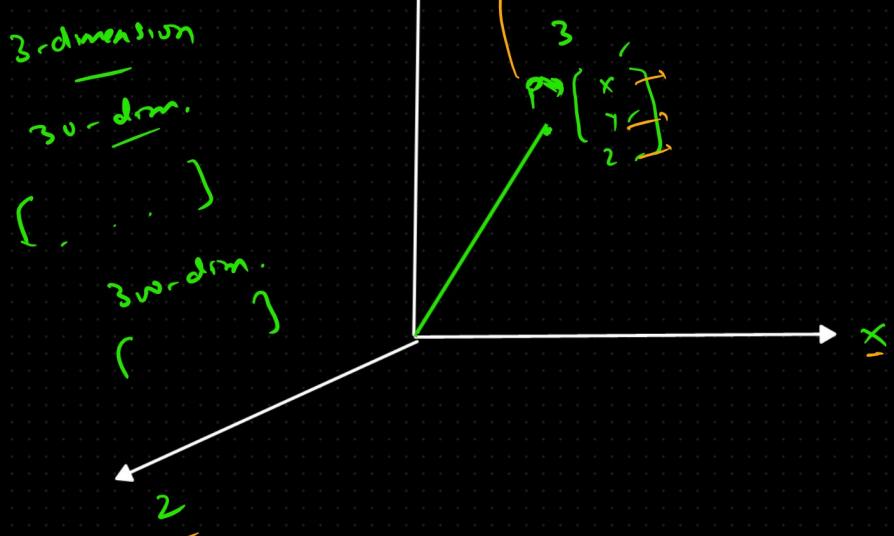
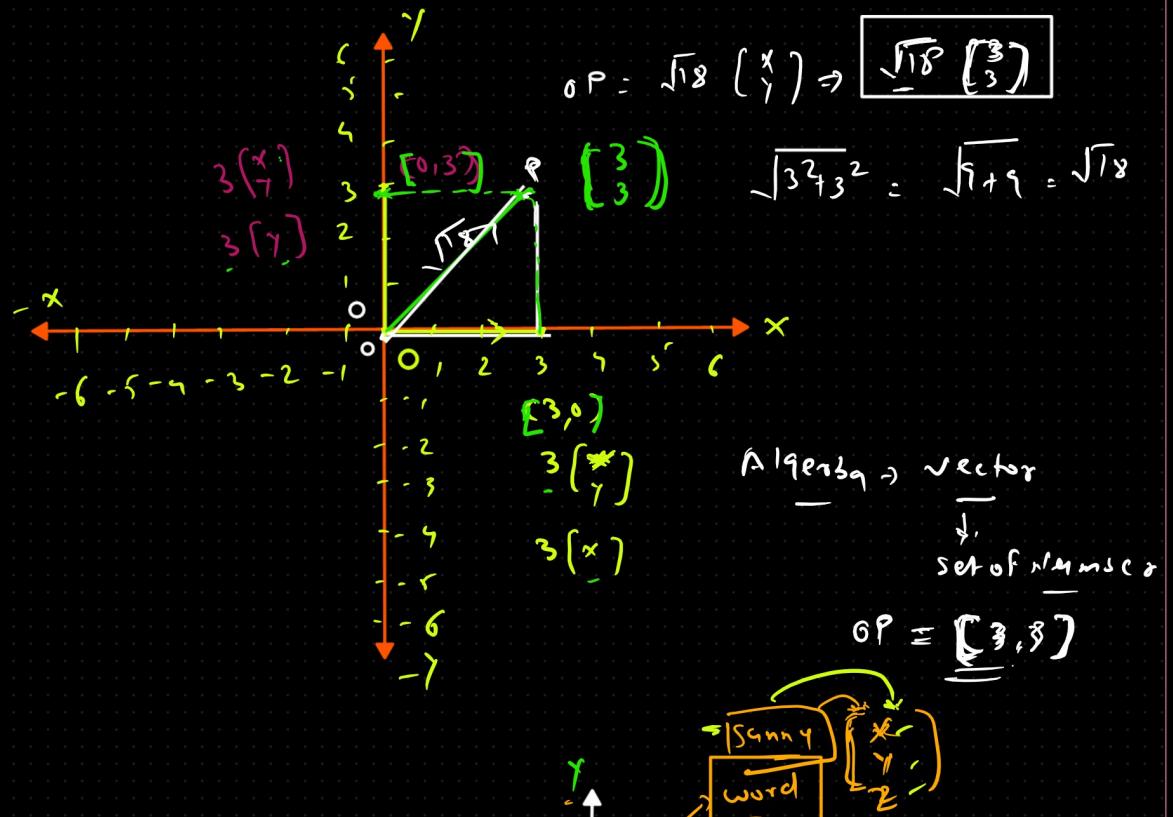
3 Sparsity → dense

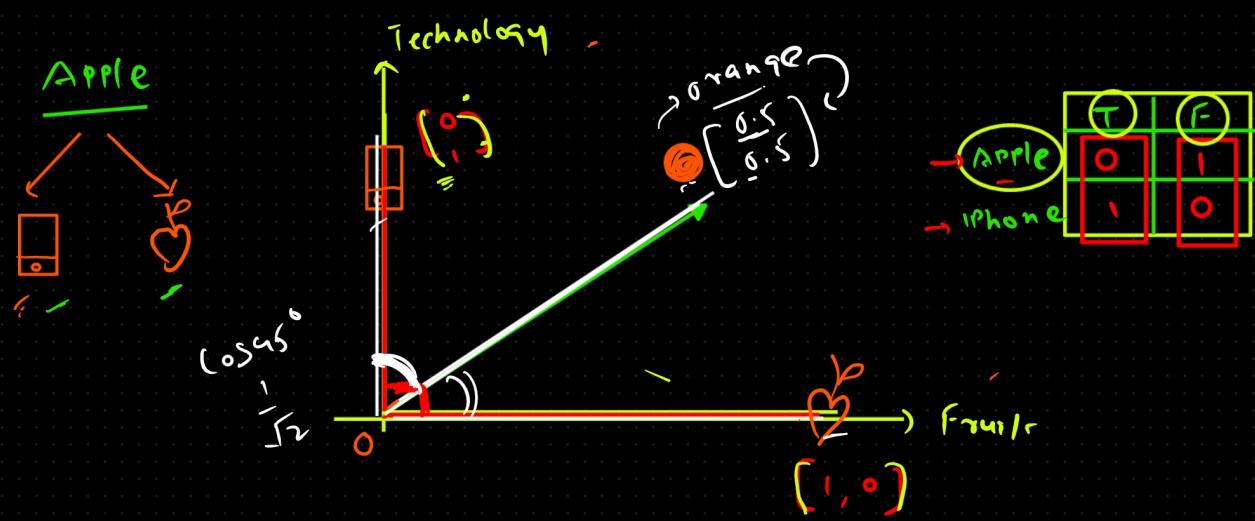
[0 0 0, 1 0 0] [0.1, 0.8, 0.9, 0.5, 1]



Vector







Similarity \Rightarrow Apple $(1, 0)$ iPhone $(0, 1)$

Apple & iPhone

$$\cos 90^\circ = 0$$

\Rightarrow ① Cosine Similarity

\Leftarrow cosine angle b/w the vector

\Rightarrow ② Dot Product

$$\text{apple} \cdot \text{iphone} \\ (\bar{x_1}, \bar{y_1}) \cdot (\bar{x_2}, \bar{y_2})$$

$$x_1x_2 + y_1y_2$$

$$(1, 0) \cdot (0, 1)$$

$$1 \times 0 + 0 \times 1$$

$$= 0$$

\Rightarrow calculate me Dot Product b/w me vector

Word2vec → Behind the scene

[0-1]

	f_1	f_2	f_3	f_4	f_5	
(1) gender	1	1	1	0.8	1	\rightarrow wordvec
King	1	1	1	0.8	1	\rightarrow S
Queen	0	0.7	0.8	0.9	1	
Man	1	0.3	0.3	0.7	1	
Woman	0	0.2	0.1	0.9	1	
Monkey	1	0	0	0.9	0	

King - Man + Woman = Queen

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0.8 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0.3 \\ 0.3 \\ 0.2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.1 \\ 0.9 \\ 0.1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.9 \\ 0.8 \\ 0.9 \\ 1 \end{pmatrix}$$

G W P W S

I have a word \Rightarrow its feature

manually

50000

Automation \Rightarrow CBOW, Skipgram

\downarrow
NN

feature \rightarrow Vocab

context = 3



data Screen

CBoW \Rightarrow Continuous Bag of words \rightarrow vocab 5

Ex:-



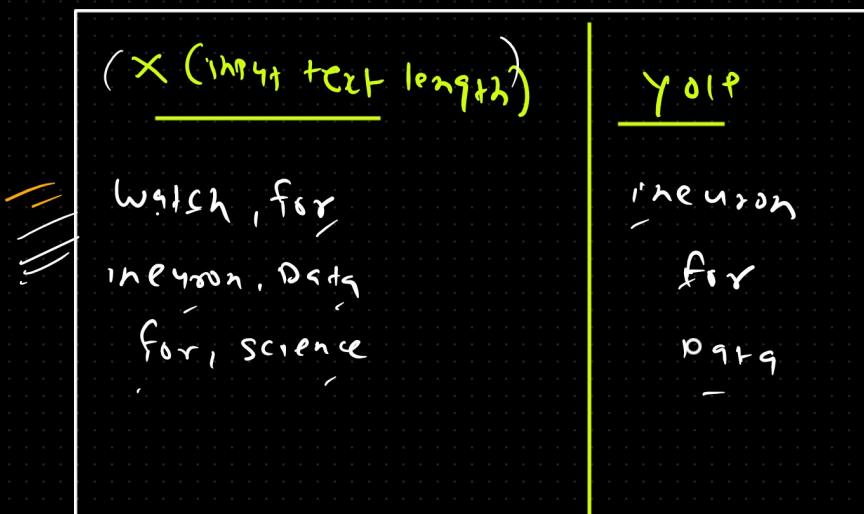
1 Context window $\Rightarrow 2, 3, 5, 7, 8, 10, \dots$
2 target

Window = 3

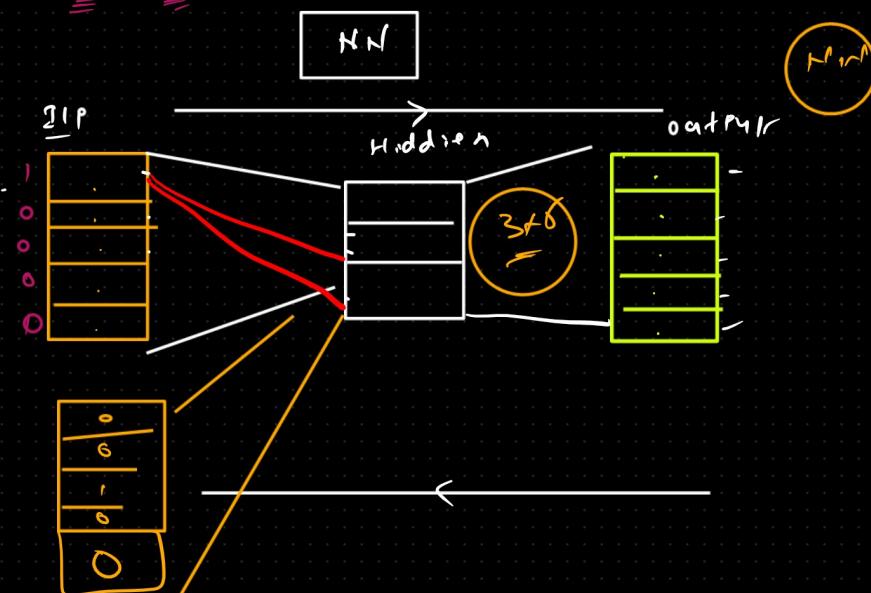
↓

Input text length

$$\text{Weight} = [0.6, 0.1, 0.9] \Rightarrow 3\text{-dim}$$



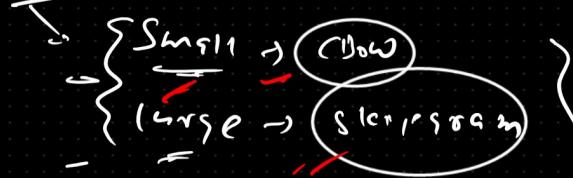
(watch, neuron, for, Data, Science)



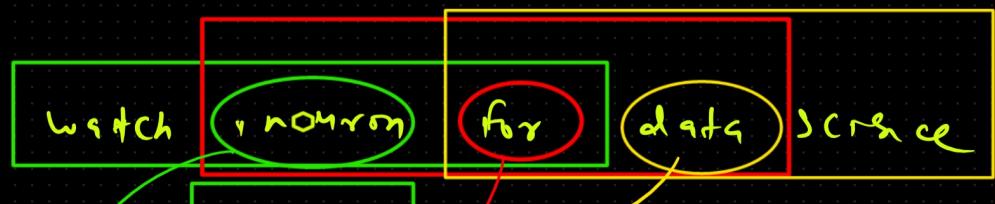
Word2vec

- 1 Data \rightarrow lots of Data
- 2 Dense (No. of Nodes) \Rightarrow Dimension
- 3 Window size
Content size
 \downarrow
Vocab

problem



skipgram \Rightarrow



Input neuron
watch, for
for, science

for

far sel.
watch, for

neuron, data

for, science

