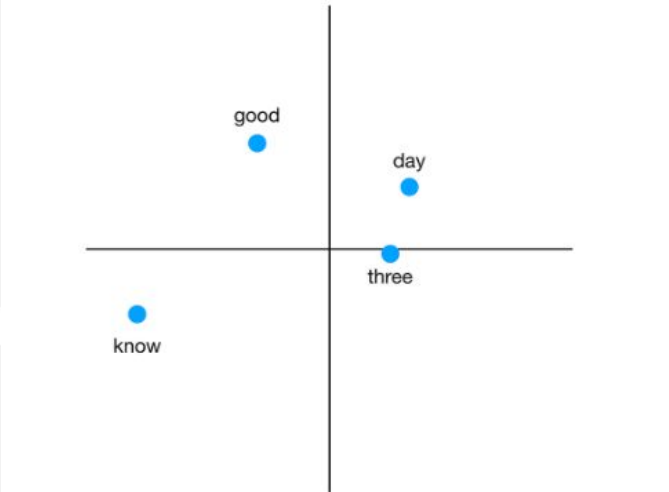

# Word vector algebra

Akshitha Kumbam

12 September, 2024

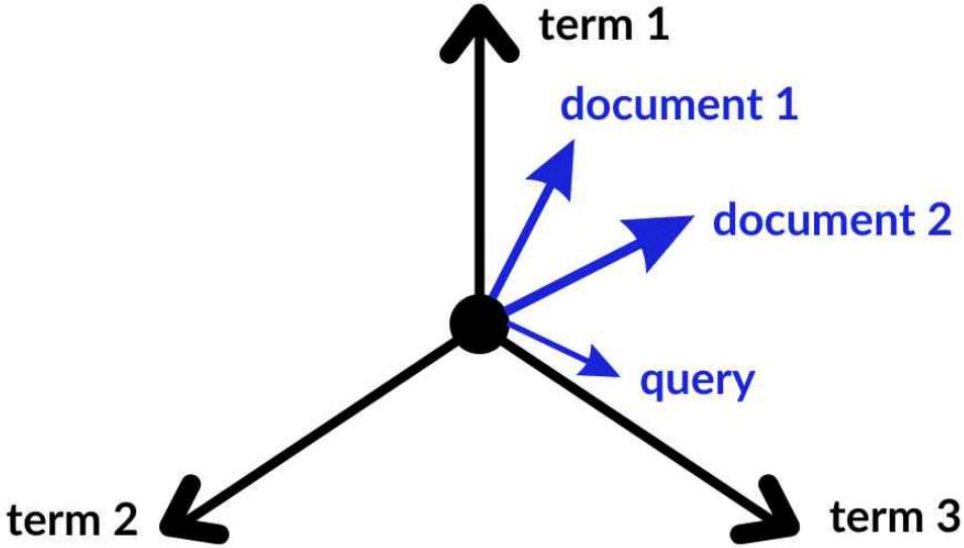
# What is word Vector Algebra?

Word vector algebra is a mathematical framework for representing words as vectors in a high-dimensional space, allowing for vector operations to capture semantic relationships.



# Vector space models

Vector space models represent words as vectors in a high-dimensional space, where semantically similar words are mapped to nearby points.




NYU

# Word Vector Algebra Operations

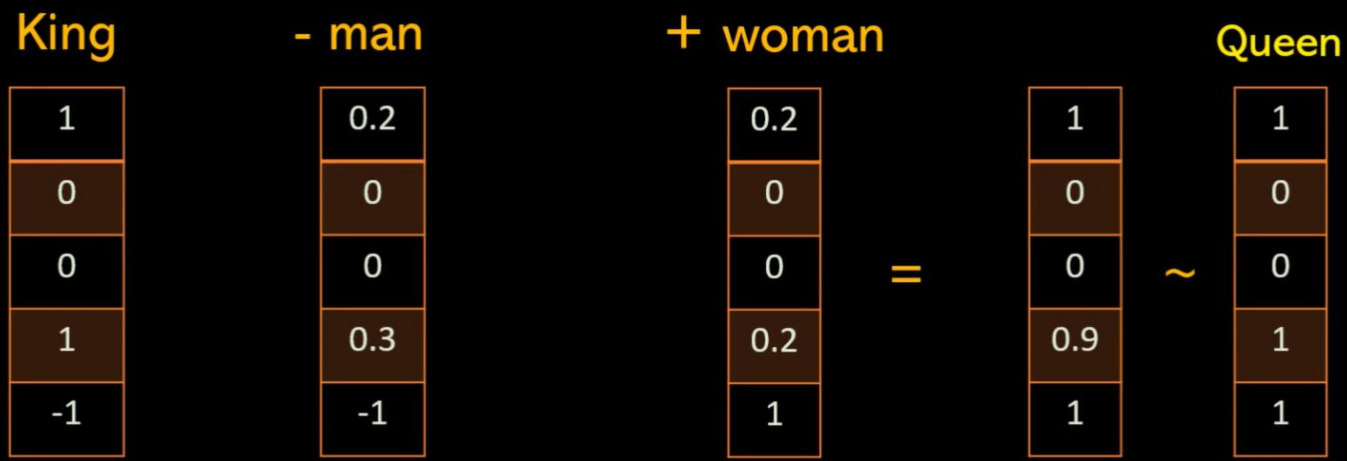
Vector addition: "big" + "house" ≈ "mansion" ; “king” + “woman” ≈ "queen"

Vector subtraction: "fast" - "slow" ≈ "speed" ; "king" - "man" ≈ "woman"

Vector multiplication: "fast" * "car" ≈ "fast car"

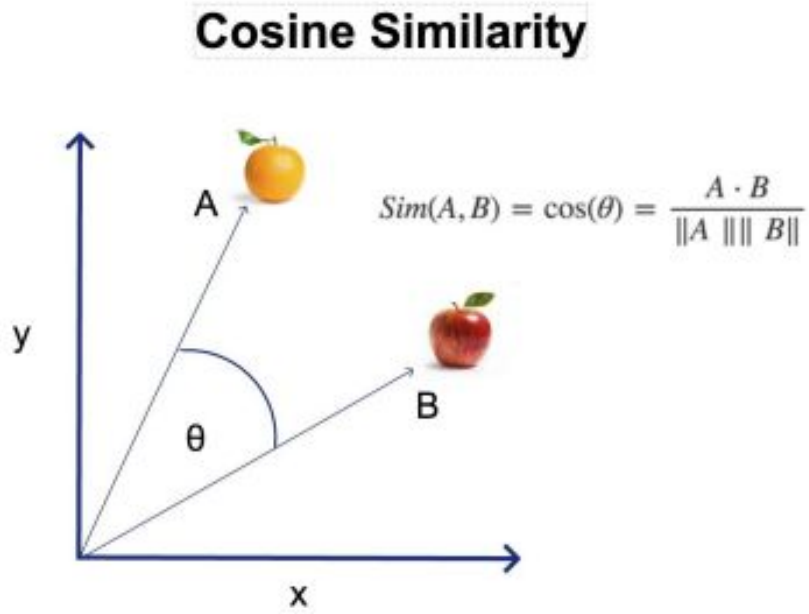
https://www.youtube.com/watch?v=hQwFeIupNP0

| | battle | horse | king | man | queen | .. | woman |
|---|---|---|---|---|---|---|---|
| authority | 0 | 0.01 | 1 | 0.2 | 1 | ... | 0.2 |
| event | 1 | 0 | 0 | 0 | 0 | ... | 0 |
| has tail? | 0 | 1 | 0 | 0 | 0 | ... | 0 |
| rich | 0 | 0.1 | 1 | 0.3 | 1 | ... | 0.2 |
| gender | 0 | 1 | -1 | -1 | 1 | ... | 1 |



# Cosine similarity

Cosine similarity measures the cosine of the angle between two vectors.




NYU

Jaccard similarity and Euclidean distance.

Cosine similarity applications -
Used for measuring semantic similarity.
For example, Sentence transformers models.

**Cosine of the angle** - Ranges from -1 to 1

# Word Analogy Task

The word analogy task involves finding a word that completes an analogy, such as 'king' is to 'man' as 'queen' is to 'woman'.

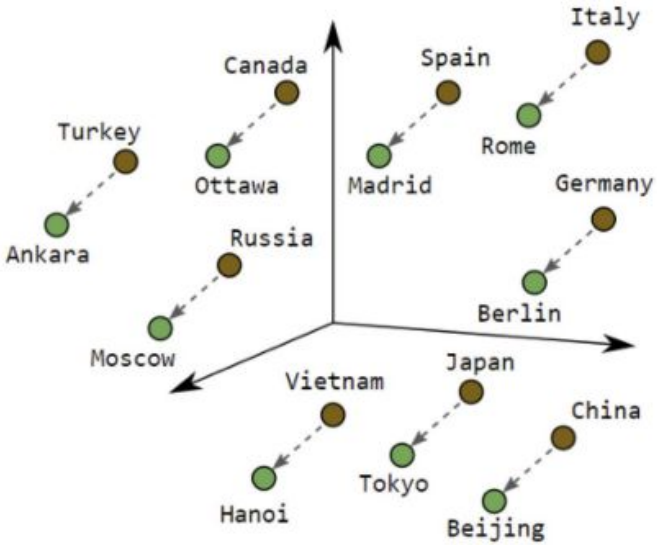
"car" is to "road" as "boat" is to "_____"

Answer: "water"

For a —> b :: c —> ?, given word vectors $v_a$, $v_b$ and $v_c$, we will find a word d such that $v_a - v_b \sim v_c - v_d$.

The difference $v_a - v_b$

represents the ‘concept’.




Country-Capital

NYU

# Bias in Word Vectors

The difference $v_a$ - $v_b$ represents the 'concept' — if a is woman and b is man, then it represents 'gender'.

Compute projections of occupations on this difference $v_a$ - $v_b$

**Extreme *she* occupations**

| | | |
|---|---|---|
| 1. homemaker | 2. nurse | 3. receptionist |
| 4. librarian | 5. socialite | 6. hairdresser |
| 7. nanny | 8. bookkeeper | 9. stylist |
| 10. housekeeper | 11. interior designer | 12. guidance counselor |

**Extreme *he* occupations**

| | | |
|---|---|---|
| 1. maestro | 2. skipper | 3. protege |
| 4. philosopher | 5. captain | 6. architect |
| 7. financier | 8. warrior | 9. broadcaster |
| 10. magician | 11. figher pilot | 12. boss |

NYU

# Word Embeddings: Word2Vec and GloVe

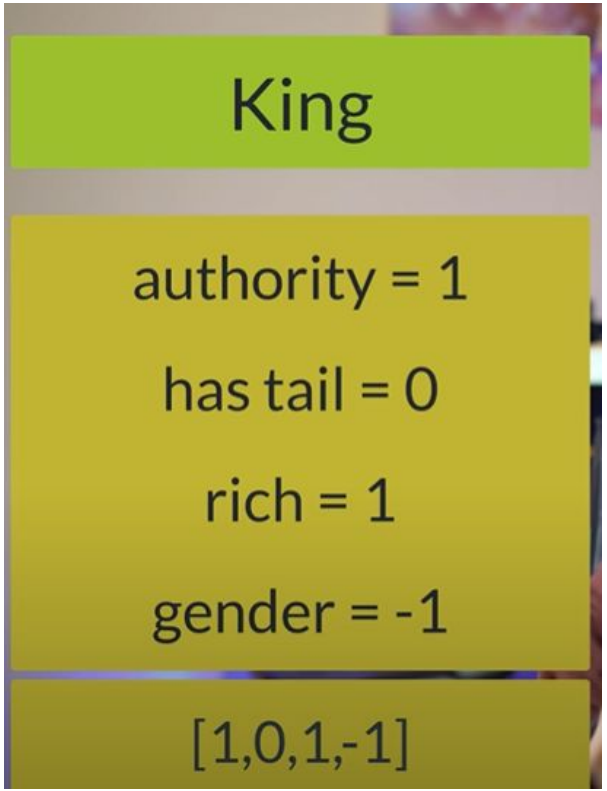
**Word2Vec** is a shallow neural network that generates word embeddings by predicting neighboring words. It uses a single layer neural network with a softmax output layer.

**Problem - Predict the missing word**

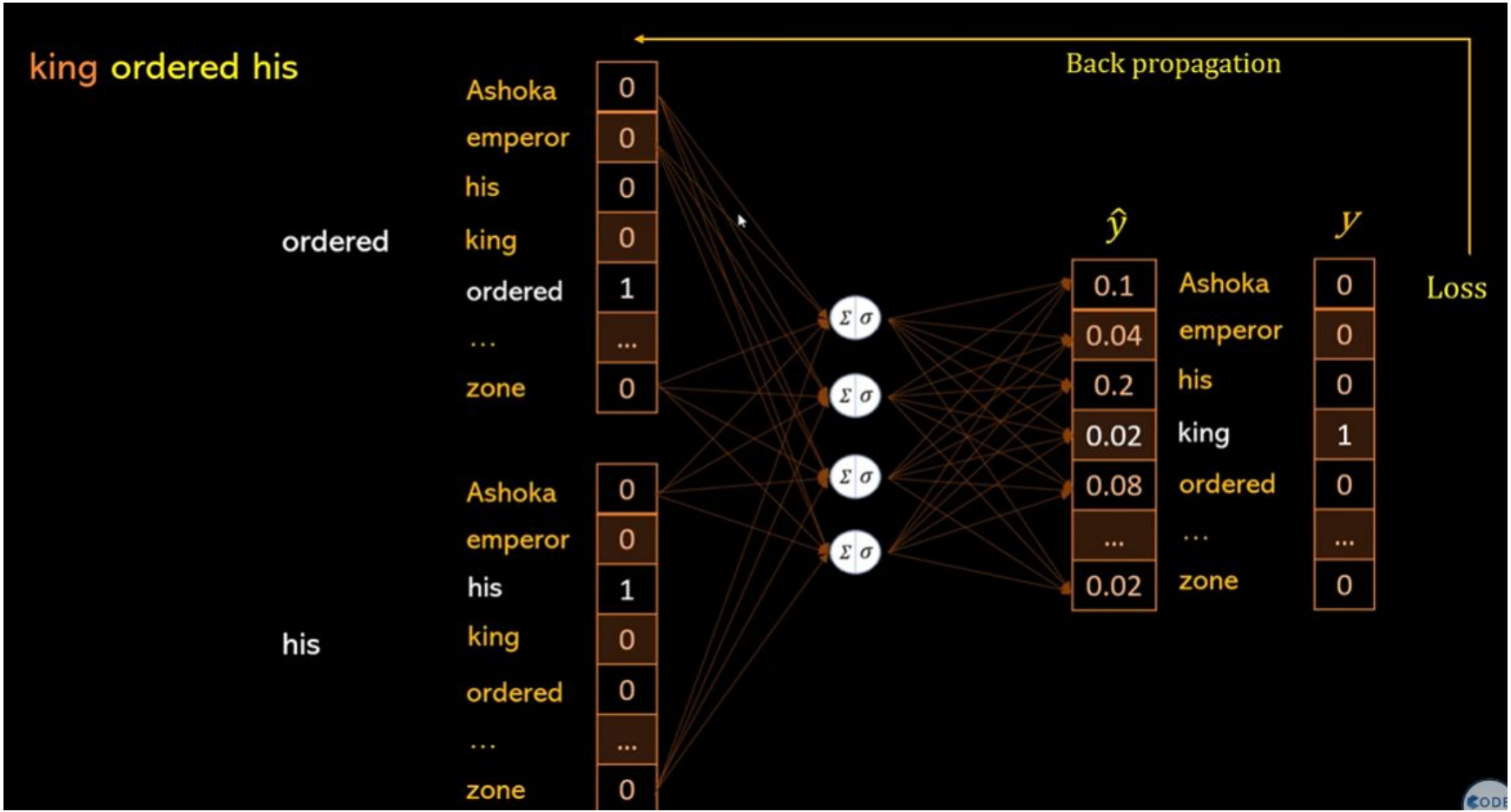
The _____ is sleeping in his palace. apple, laptop, king, whale

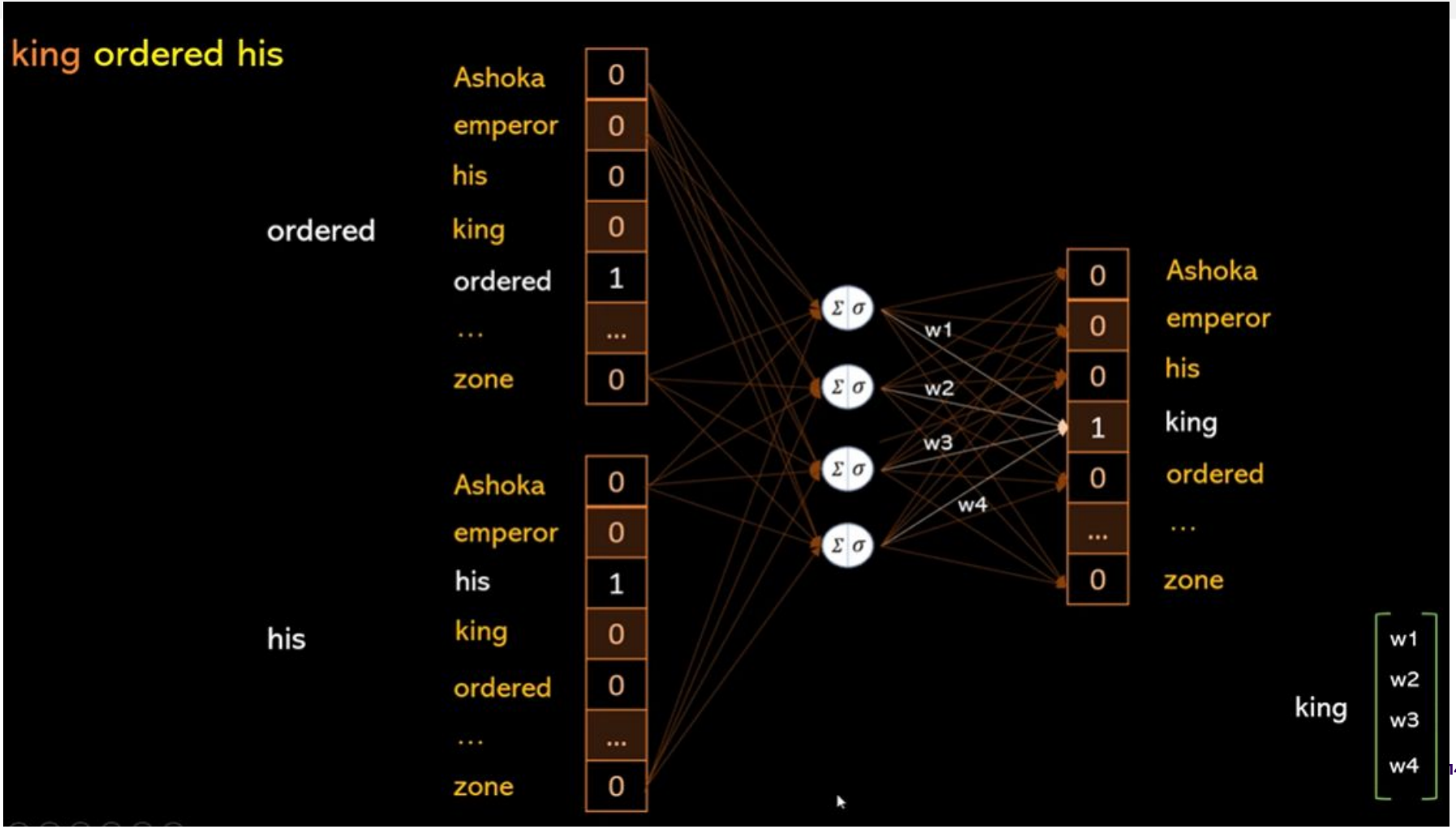
**Result - Word embeddings**

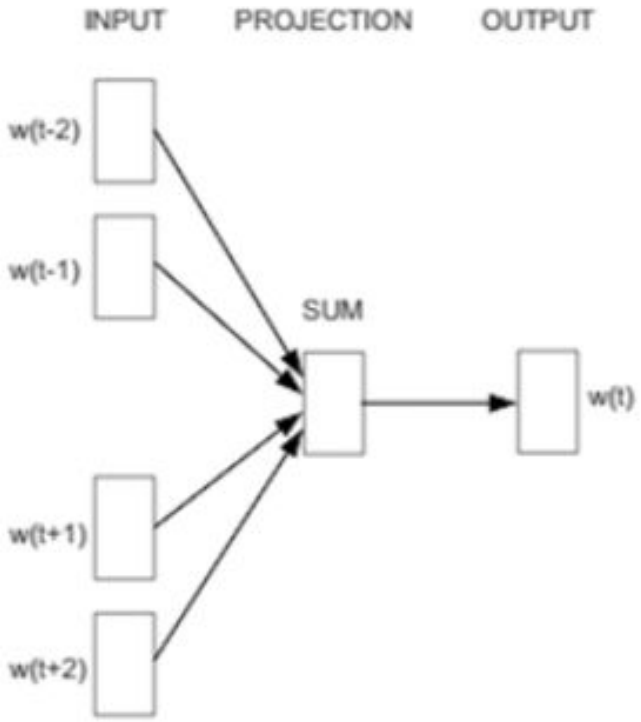
NYU


King
authority = 1
has tail = 0
rich = 1
gender = -1
[1,0,1,-1]

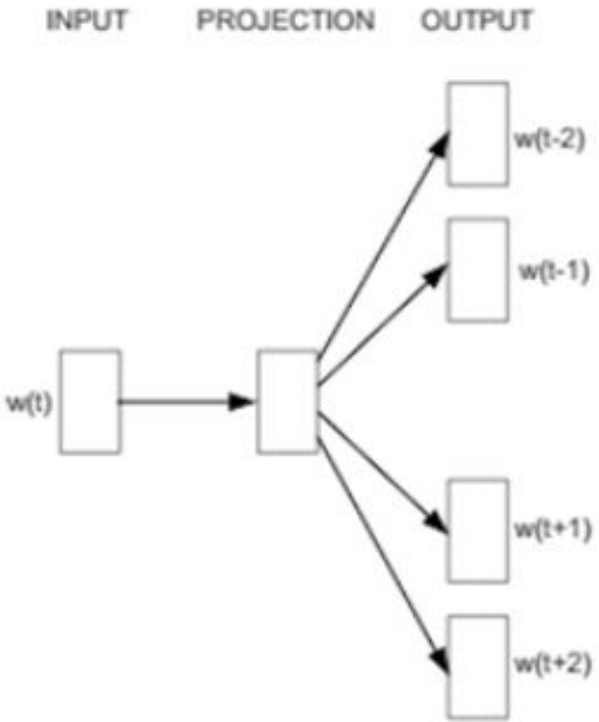
Word2Vec uses the neighboring words to predict the missing word.




king ordered his
ordered
Ashoka
emperor
his
king
ordered
…
zone
0
0
0
0
1
…
0
his
Ashoka
emperor
his
king
ordered
…
zone
0
0
1
0
0
…
0
w1
w2
w3
w4
0
0
0
1
0
…
0
Ashoka
emperor
his
king
ordered
…
zone
king
w1
w2
w3
w4




CBOW




Skip-gram

NYU

The model uses a self-supervised method of learning and represents words as dense vectors in a high-dimensional space.

# Drawbacks of Word2Vec

Word2Vec relies solely on local information, meaning that the semantics learned for a target word only depend on the surrounding context words. This limitation can lead to:

Contextual ambiguity

Lack of global understanding

Difficulty with rare words

# GloVe

GloVe is a count-based method that represents words as vectors based on their co-occurrence patterns. Derives the semantic relationship between words using **word-word co-occurrence matrix.**

1. **I love NLP**
2. **I love to make videos**

| | I | love | NLP | to | make | videos | . |
|---|---|---|---|---|---|---|---|
| I | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| love | 2 | 0 | 1 | 1 | 0 | 0 | 0 |
| NLP | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| to | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| make | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| videos | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| . | 0 | 0 | 1 | 0 | 0 | 1 | 0 |

$w \in \mathbb{R}^d$ are word vectors

probe word

$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

co-relations between the word w*i* and w*j*

co-occurrence probabilities for the word w*j* and w*k*

$X_{ij}$ tabulate the number of times word $j$ occurs in the context of word $i$.

$X_i = \sum_k X_{ik}$

$P_{ij} = P(j|i) = X_{ij}/X_i$

# Drawback of Glove

Memory intensive process: as a faster training process we need to keep a co-occurrence matrix in RAM as a hash map and perform co-occurrence increments.

# Further,

FastText (2016)

BERT (2018)

RoBERTa  (2019)

WordPiece Embeddings (2016)

DistilBERT (2019)

Sentence embeddings :

Sentence-BERT

NYU

# Summary

Word embeddings are a fundamental component of natural language processing (NLP) that enable machines to understand the meaning of words.

From the early days of Word2Vec and GloVe to the latest transformer-based models, word embeddings have evolved significantly.

Contextualized word embeddings like BERT, RoBERTa, and XLNet have revolutionized the field by capturing context-dependent word meanings.

Meanwhile, sentence embeddings like Sentence-BERT have enabled machines to understand the meaning of entire sentences.