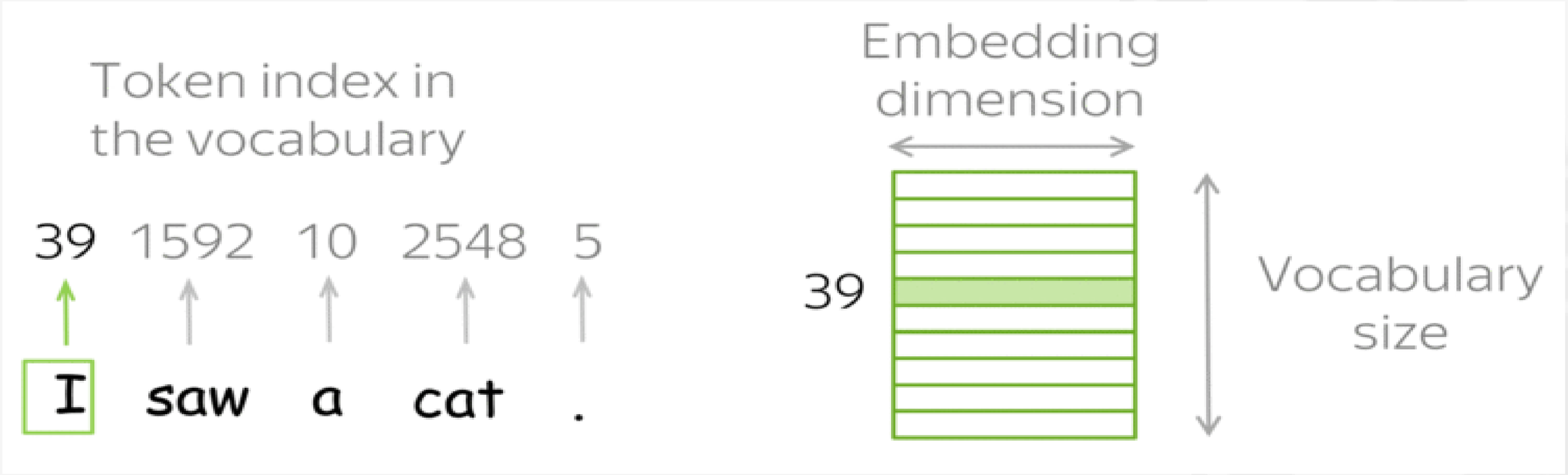
NLP WORKSHOPS WITH LLM INSIGHT

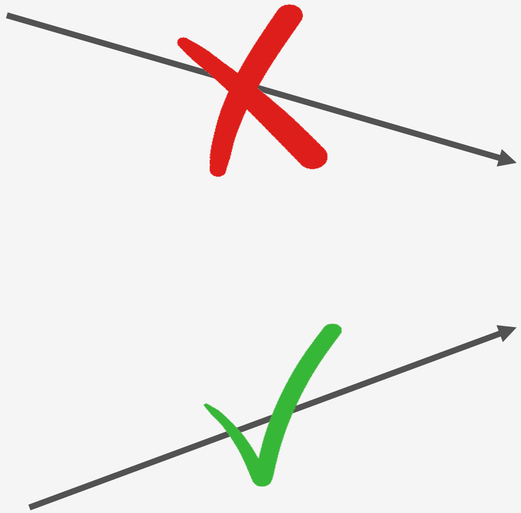
INTRODUCTION TO EMBEDDINGS



Mathematical Representation of words

WHY TEXT EMBEDDING

“so do all who live to see such times.”



ML MODEL

0	0	8	0
2	25	0	7
13	0	12	9
15	32	10	0

WHY TEXT EMBEDDING

“so do all who live to see such times.”

0	0	8	0
2	25	0	7
13	0	12	9
15	32	10	0

WHY TEXT EMBEDDING

- One-hot encoding
- Count based representation
- Embeddings

ONE-HOT ENCODING

Vocabulary = (I Apple / You Cat / Bag Dog)

ONE-HOT ENCODING

Vocabulary = (I, You, Bag, Apple, Cat, Dog)

I	You	Bag	Apple	Cat	Dog
0	0	1	0	0	0
0	1	2	3	4	5

BAG OF WORDS

A wizard is never late.

He arrives precisely when he means to.

BAG OF WORDS

A wizard is never late.

A wizard	wizard is	is never	never late	he arrives	…
1	1	1	0	1	…

He arrives precisely when he means to.

0	0	0	0	1	…

TF-IDF

$$TF = \frac{\textit{number of times this words occurs}}{\textit{number of words in the documents}}$$

$$IDF = \log \frac{\textit{number of documents}}{\textit{number of documents where this word occurred}}$$

$$TFIDF = TF * IDF + 1$$

A wizard is never late, nor is he early.

COUNT-BASED REPRESENTATION

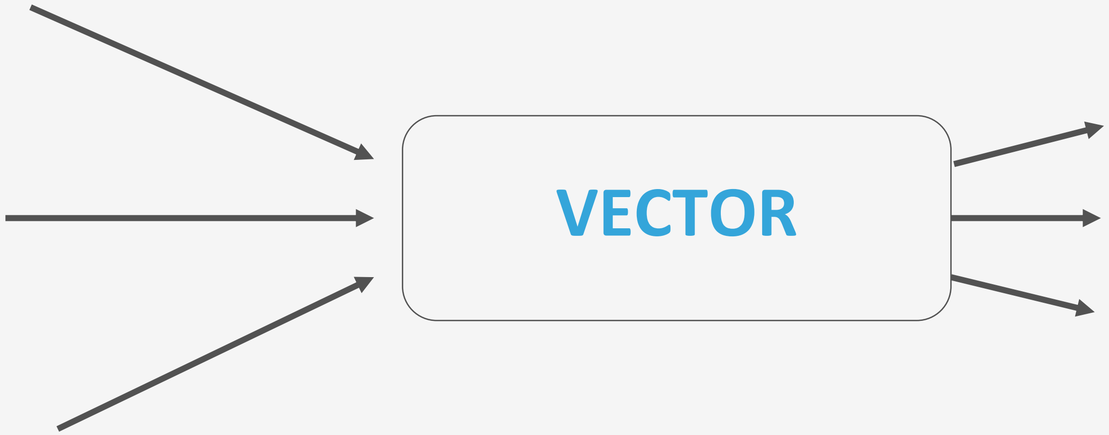
- No context
- Unknown words
- Sparse vectors

EMBEDDINGS

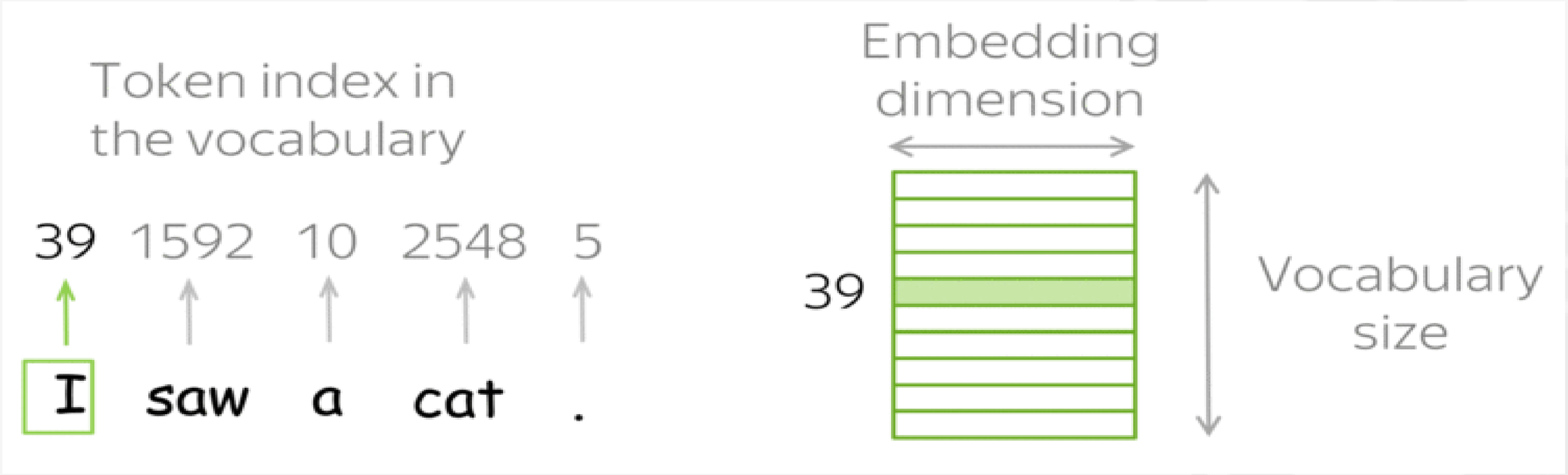
I

Love

Basketball



WHAT IS DENSE VECTOR



WHAT IS DENSE VECTOR

Tea

Coffee

Breakfast

Enjoy

Drink

WHAT IS DENSE VECTOR

Tea

Pea

Breakfast

Enjoy

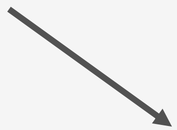
Drink

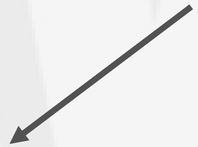
WHAT IS DENSE VECTOR

Tea

Pea

Coffee





Distance
0.3

Distance
0.7

WHAT IS EMBEDDINGS SPACE

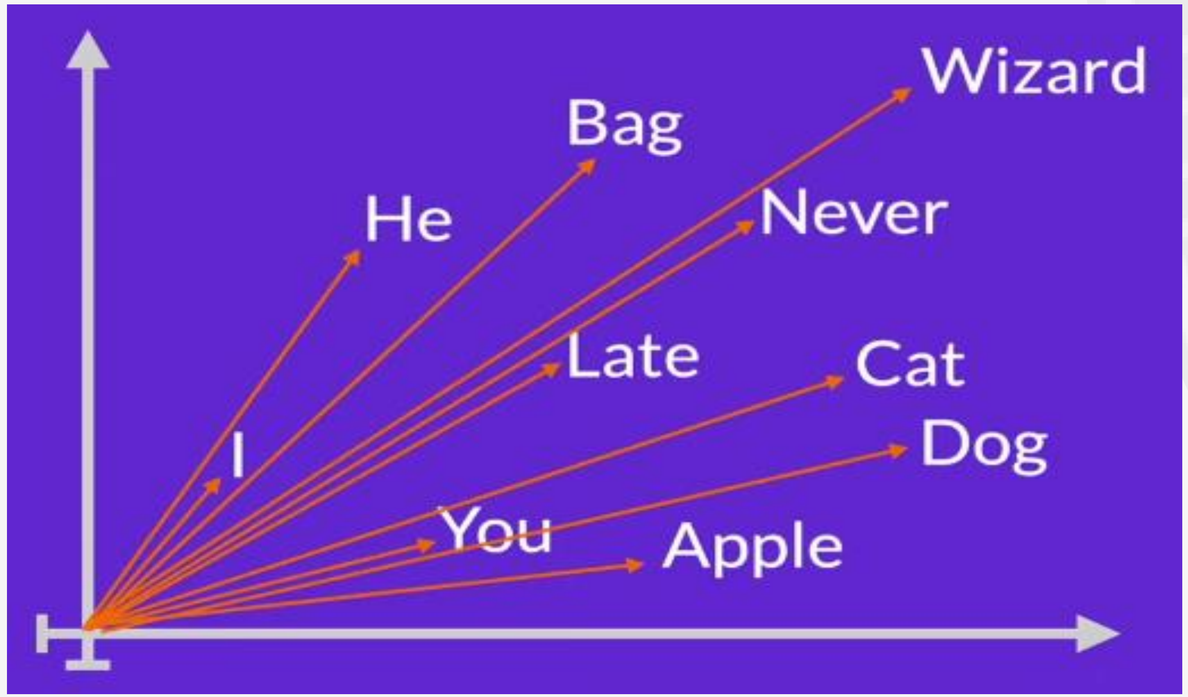
I	You	He	Late	Never	Bag	Wizard	Apple	Cat	Dog
2	11	3	7	13	5	18	8	20	23

Similarity

18 – 8 = 10

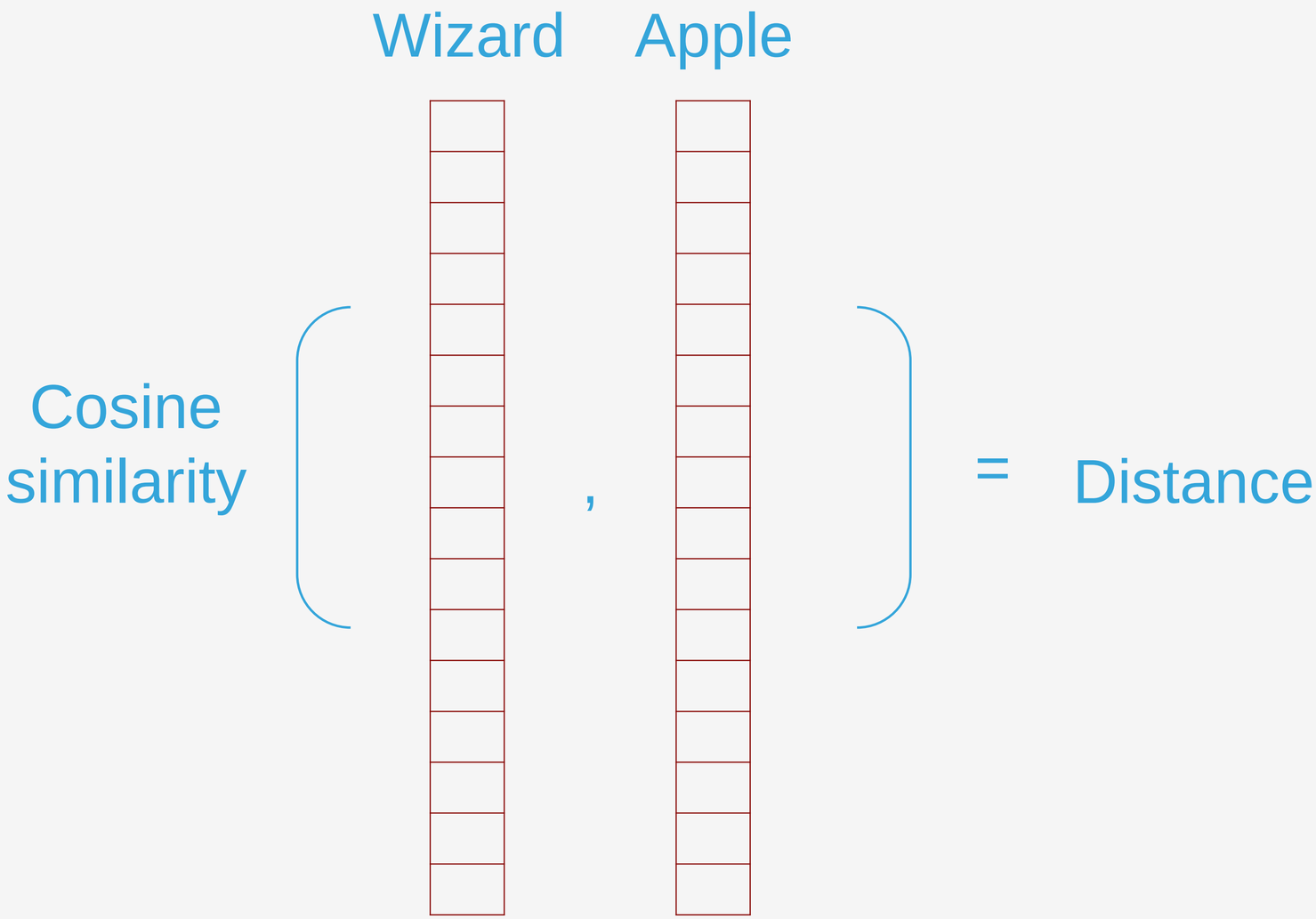
WHAT IS EMBEDDINGS SPACE

I	You	He	Late	Never	Bag	Wizard	Apple	Cat	Dog
[2,5]	[11,3]	[3,8]	[7,5]	[13,10]	[5,8]	[18,20]	[8,4]	[20,2]	[23,1]

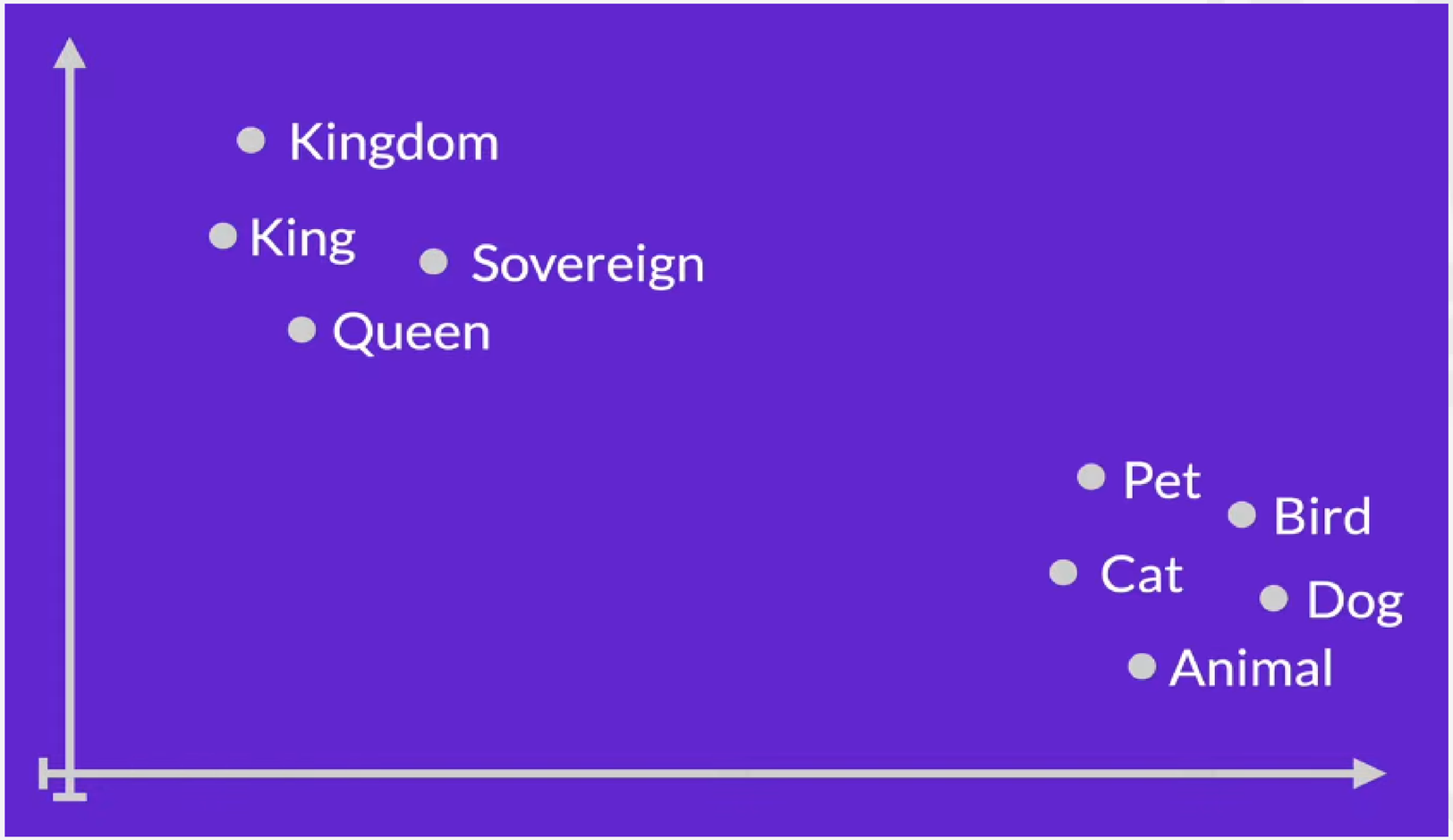


WHAT IS EMBEDDINGS SPACE

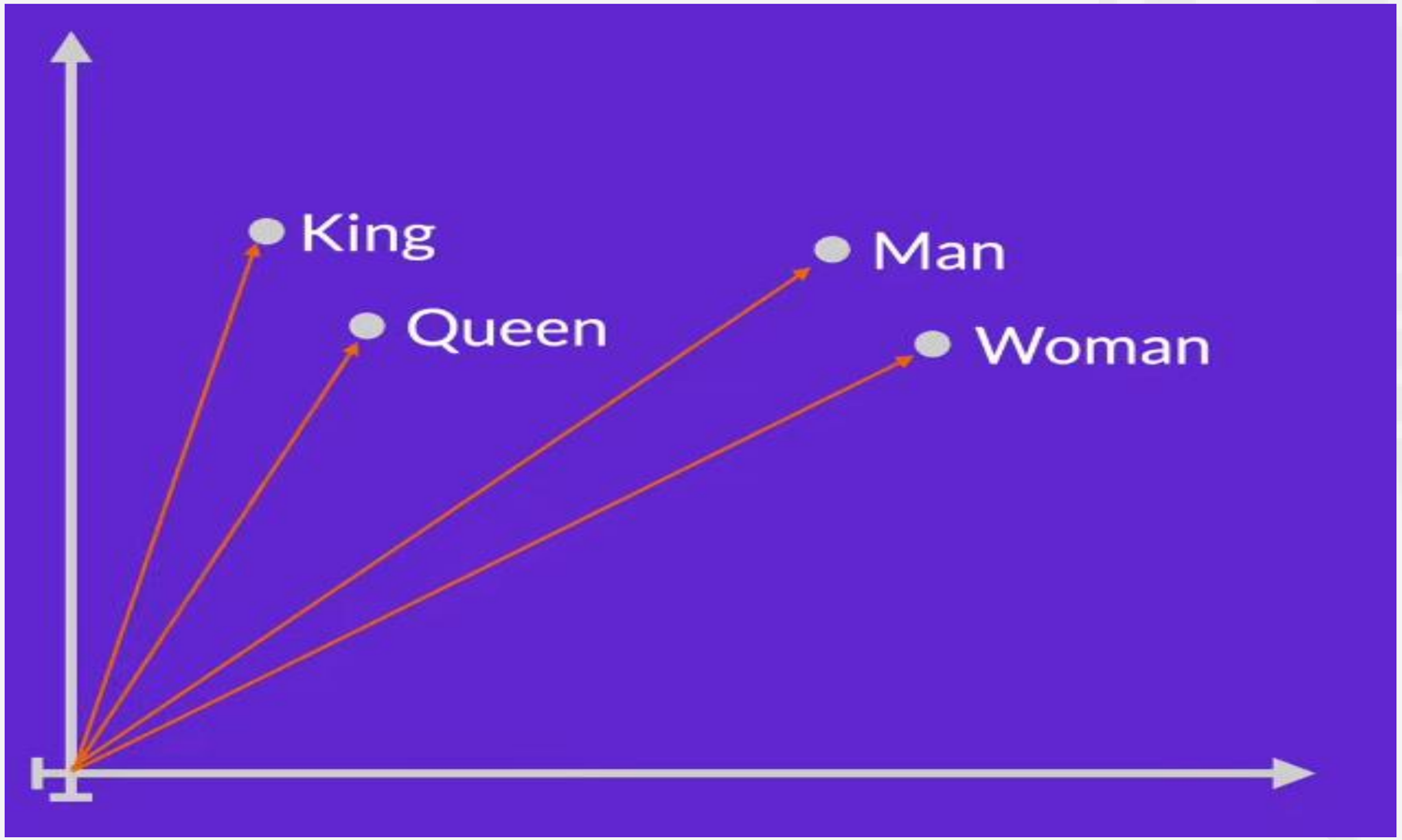
WHAT IS EMBEDDINGS SPACE



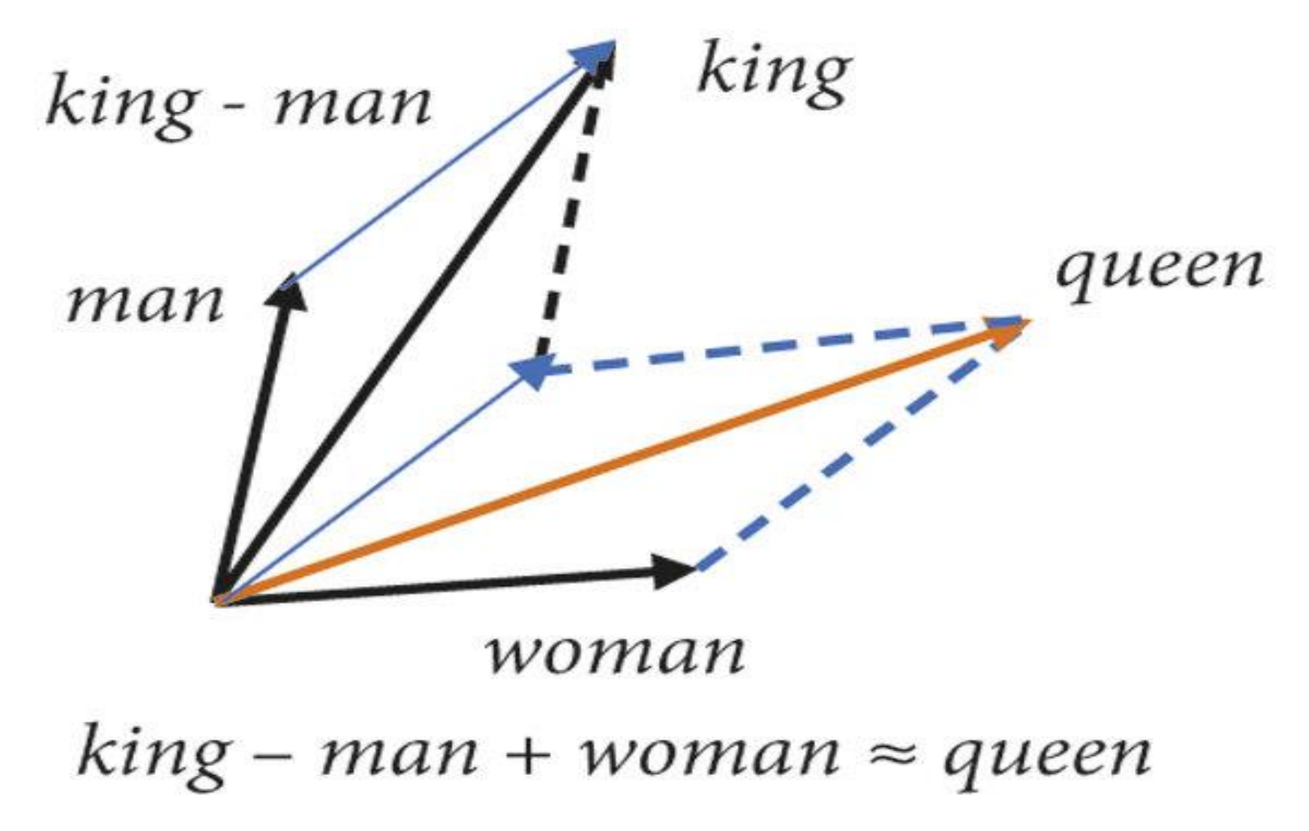
WHAT IS EMBEDDINGS SPACE



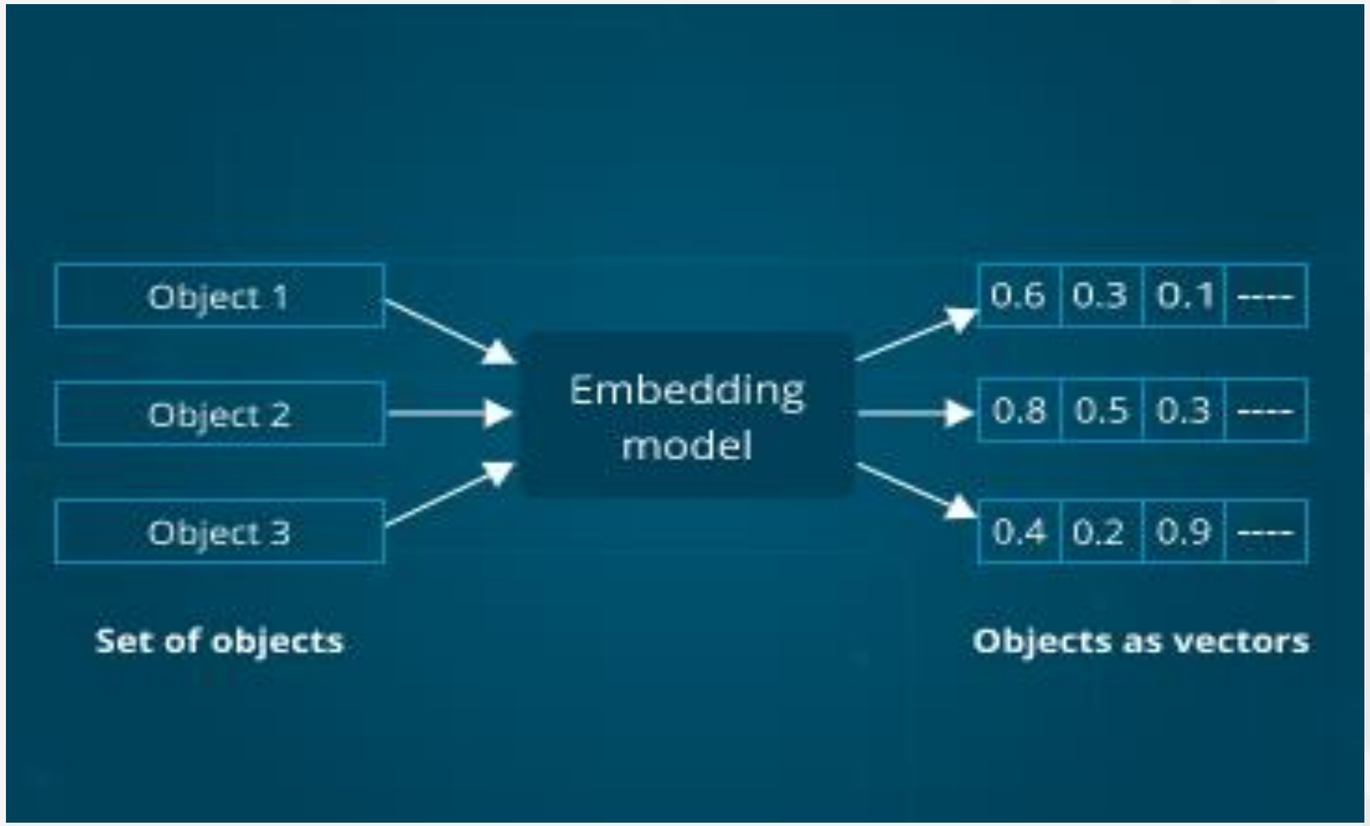
WHAT IS EMBEDDINGS SPACE



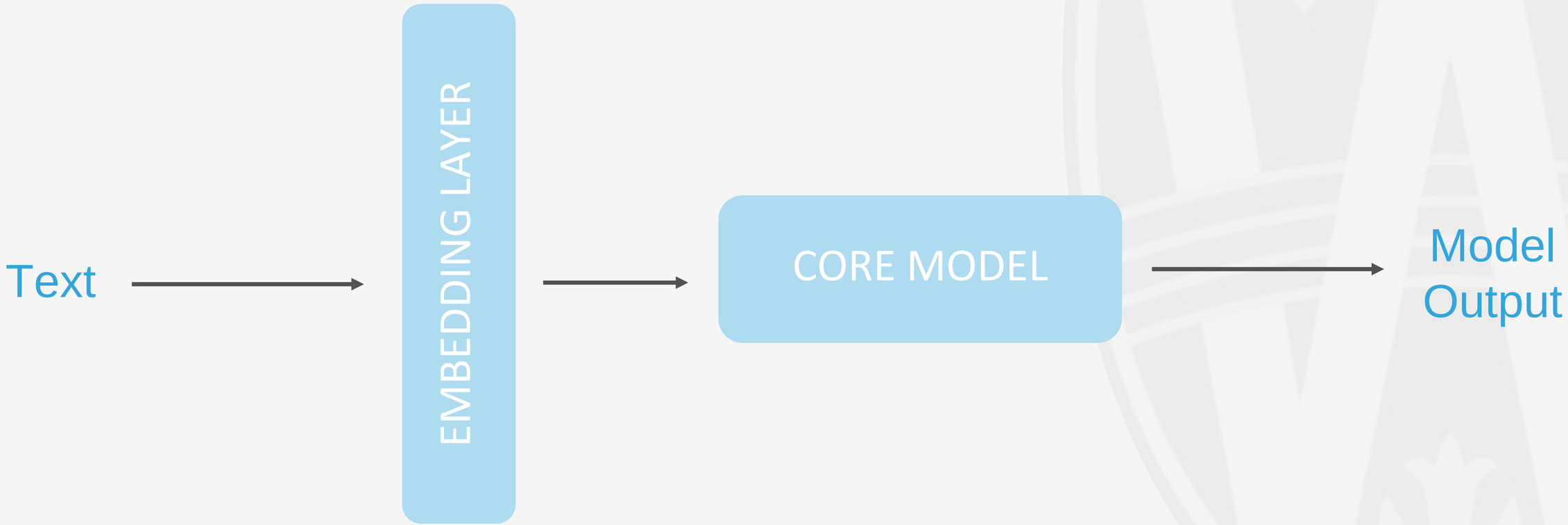
WHAT IS EMBEDDINGS SPACE



HOW ARE WORD EMBEDDINGS MADE



HOW ARE WORD EMBEDDINGS MADE



HOW ARE WORD EMBEDDINGS MADE

- **Pros:**

+ specialized embedding model

- **Cons:**

- Huge training set


One-hot
Encoded word
WORD2VEC
Word
Embedding

… I wish none of this has happened. So do all who live to see such times. But that is not up to us to decide. All we have to decide is what to do with the time that is given to us…

GLOBAL VECTORS (GLOVES)

Local Dependencies + Global Context

Word2vec

	Tea	Coffee	Pea	Apple	Monkey	App
Tea	1	6	0	2	1	0
	6	1	0	1	1	1
Pea	0	0	1	0	2	0
Apple	2	1	0	1	2	0
Monkey	1	1	2	2	1	1
App	0	1	0	0	1	1

GLOBAL VECTORS (GLOVES)

Tea

Coffee

Tea ● Coffee = log (Likelihood of tea and coffee being in the same sentence)


RPM Feedback - Zahra