Beginner's Guide to Vector Databases

AI by Hand ✍️

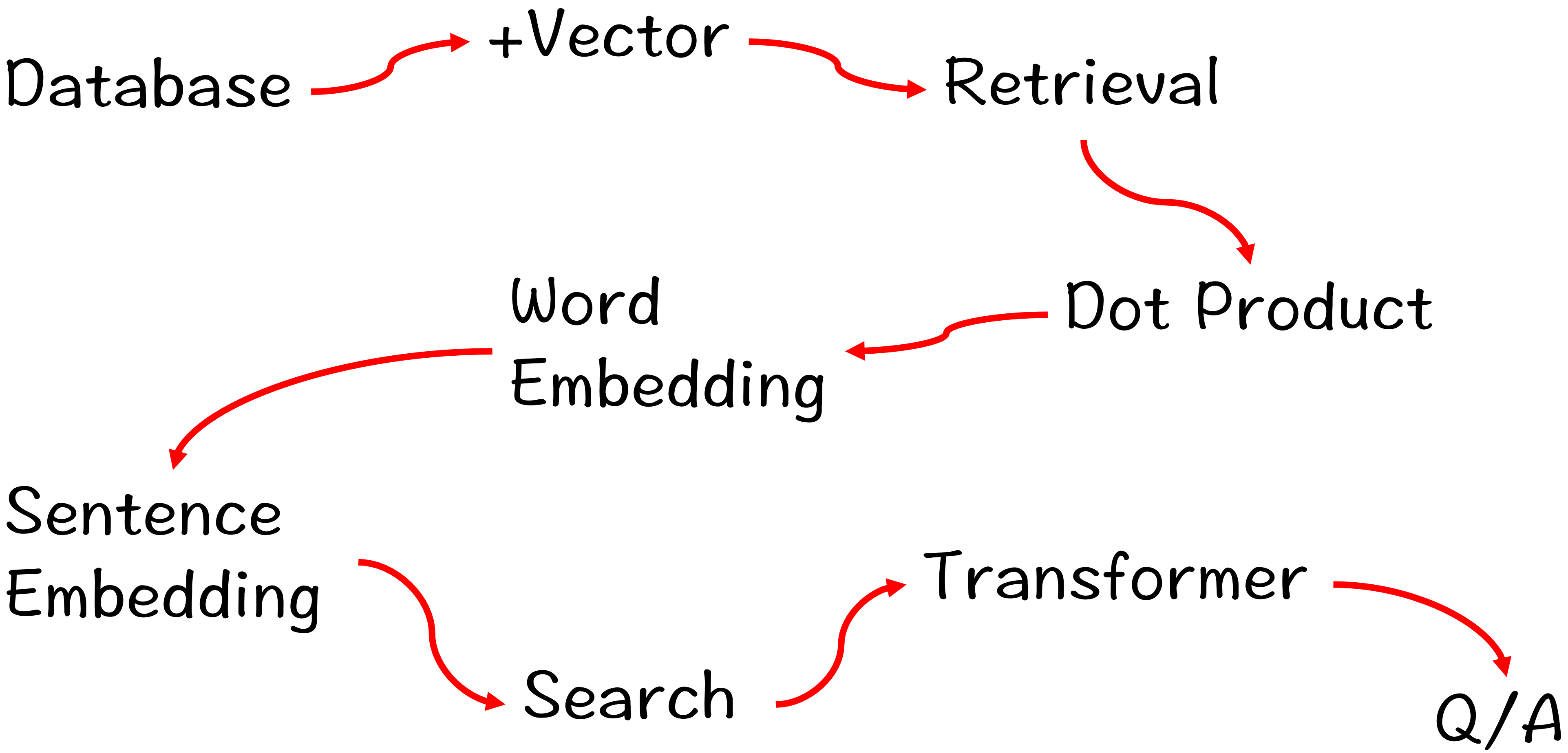
Prof. Tom Yeh

Hosted by:

SingleStore

University of Colorado Boulder

Roadmap



Database

+Vector

Retrieval

Dot Product

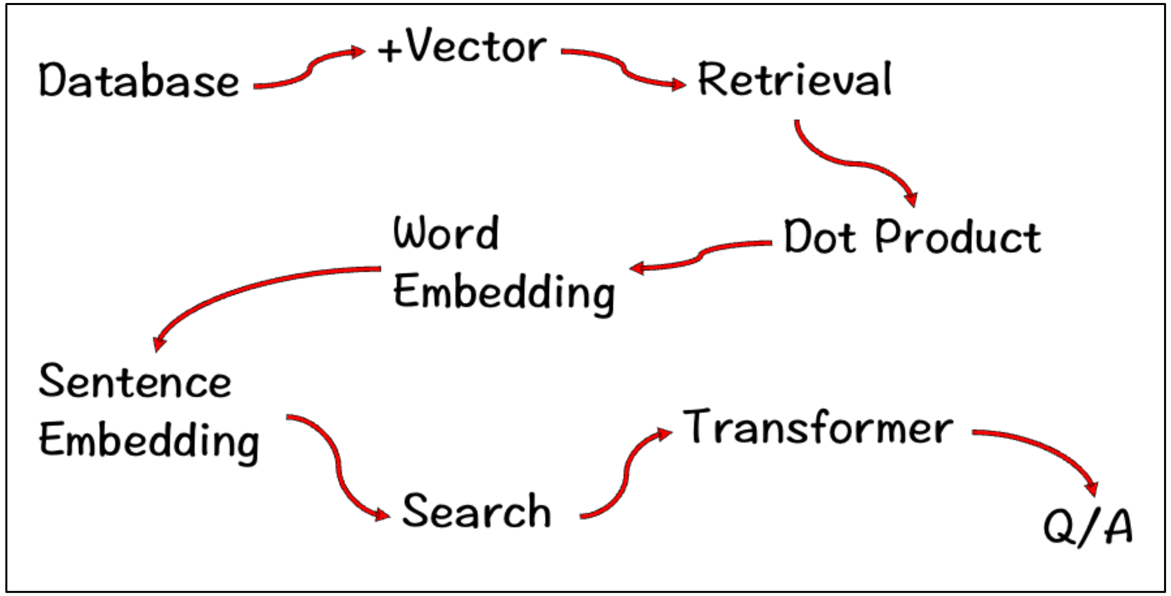
Word Embedding

Sentence Embedding

Search

Transformer

Q/A



Database
+Vector
Retrieval
Dot Product
Word Embedding
Sentence Embedding
Search
Transformer
Q/A

Database

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado
Boulder

Fun fact

There are ________900________ millions dogs in the world!

How to create a table?

SQL: CREATE TABLE animals

(id INT,

name VARCHAR(10),

size INT,

pop INT)

id	name	size	pop

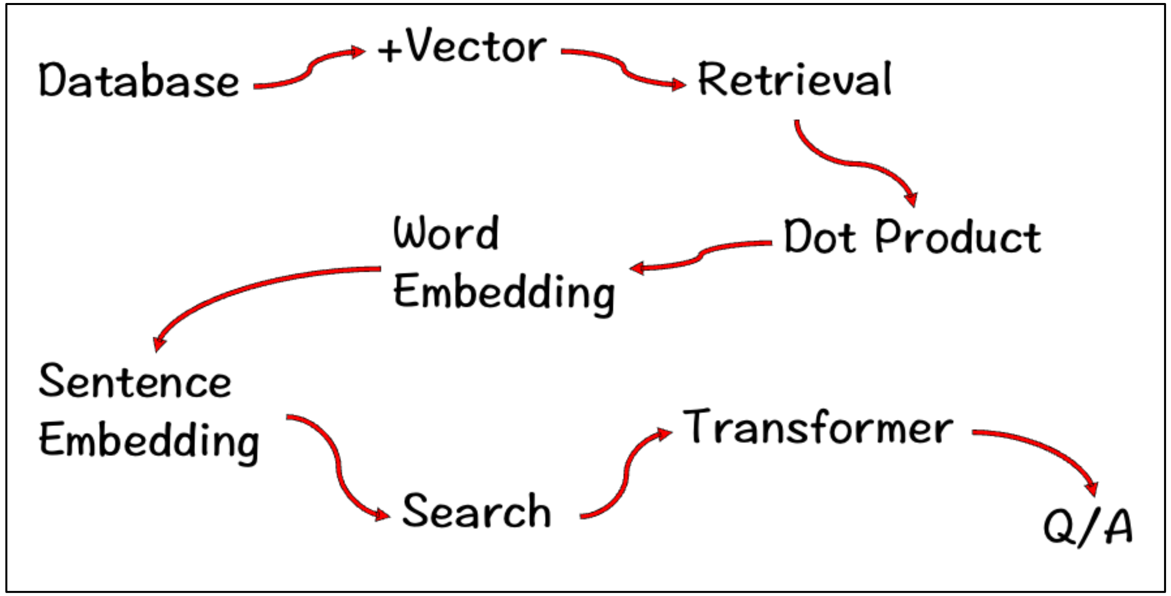
How to insert a record?

SQL:

INSERT ________ INTO animals

VALUES ________ (1, dog, 2, 900)

id	name	size	pop
1	dog	2	900



Database → +Vector → Retrieval → Dot Product → Word Embedding → Sentence Embedding → Search → Transformer → Q/A

Vector Database

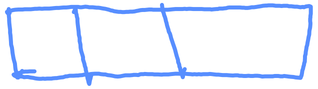
Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado
Boulder

How to create a vector database?

SQL:

```
CREATE TABLE animals
( id INT,
  name VARCHAR(10),
   size INT,
   pop INT,
   emb VECTOR(3) not null )
```

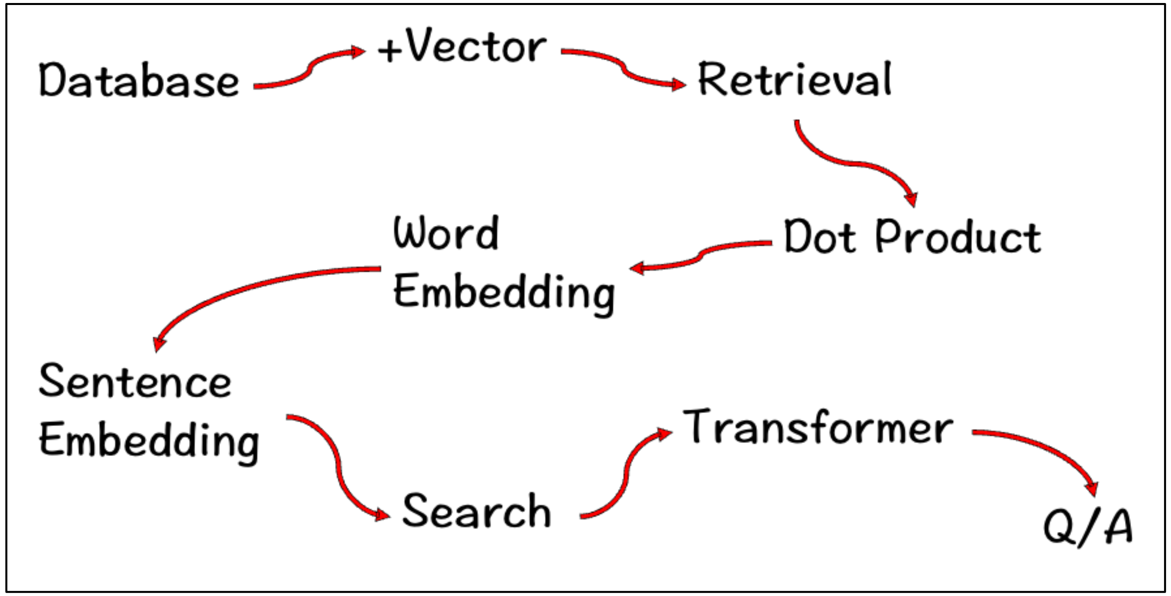
id	name	size	pop	emb
				

How to insert a record with a vector?

SQL:

```
INSERT INTO animals
VAUES (1, dog, 2, 900, [2, 1, 0])
```

id	name	size	pop	emb
1	dog	2	900	2 1 0



Database

+Vector

Retrieval

Dot Product

Word Embedding

Sentence Embedding

Search

Transformer

Q/A

Retrieval

Beginner's Guide to Vector Databases - AI by Hand ✍

University of Colorado
Boulder

Which record is relevant to the query “cat”?

Query

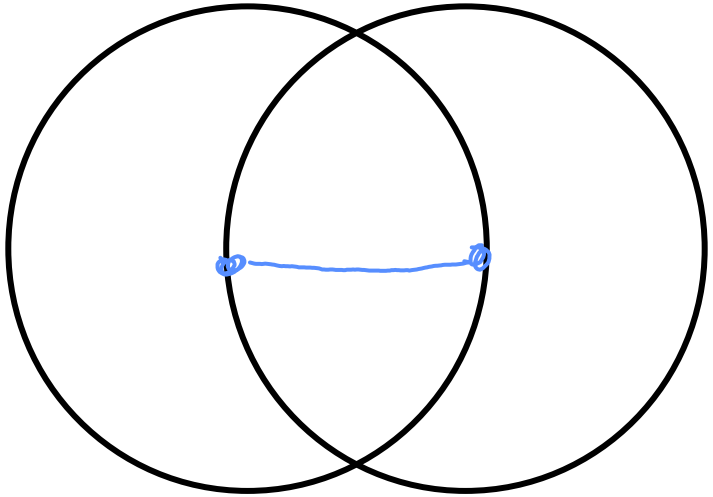
cat

1	2	0

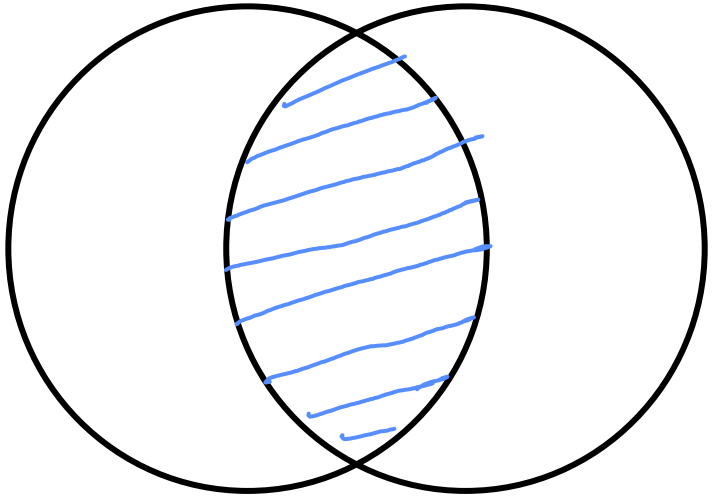
id	name	size	pop	emb
1	dog	2	900	2 1 0
2	bat	1	10000	0 1 2

Draw distance vs similarity

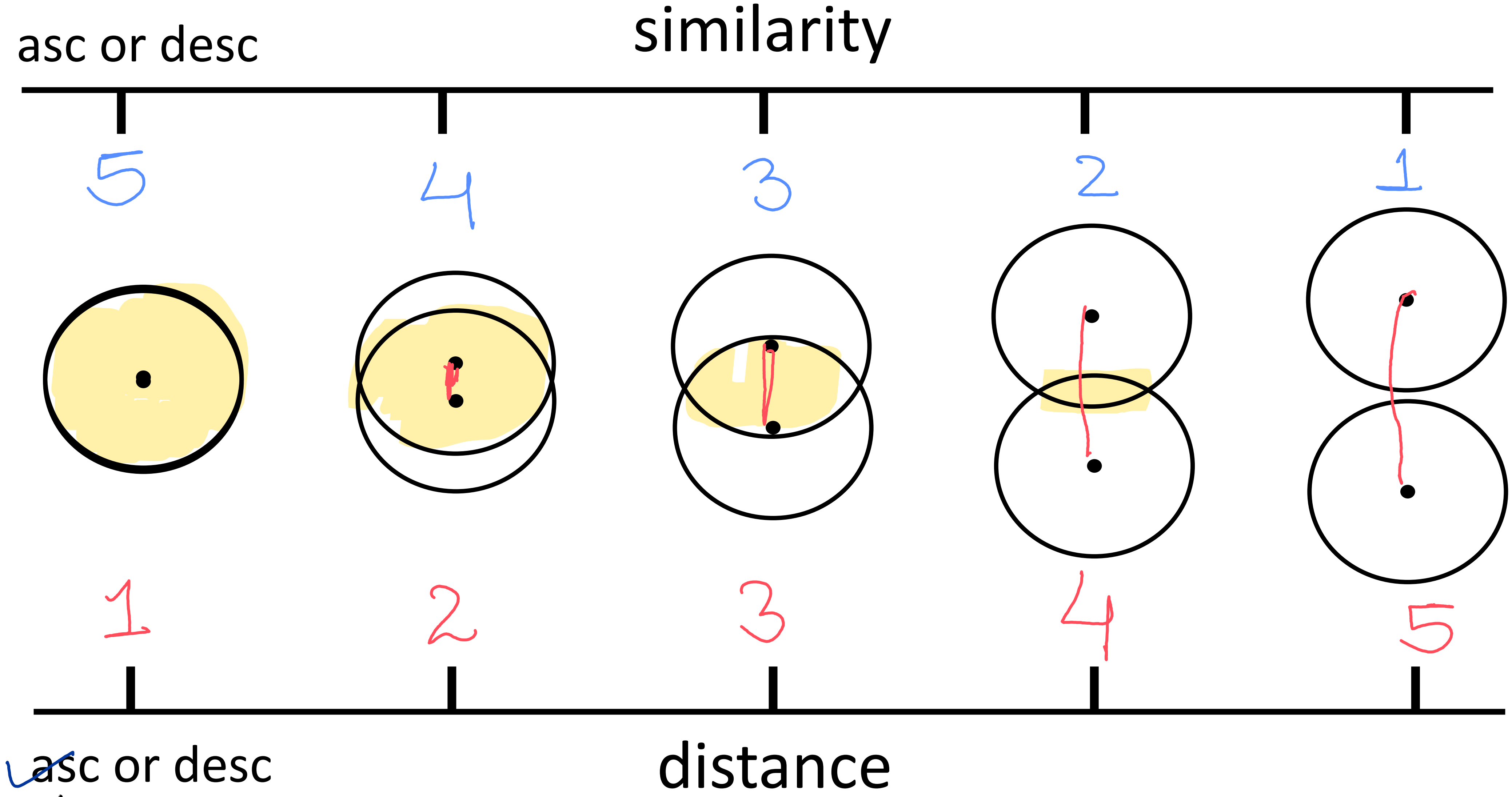
distance



similarity



Distance vs similarity on a scale of 1 to 5



asc or desc

similarity

asc or desc

distance

How to retrieve by similarity? (dot product)

↳ descending

SELECT name, emb<____>[1, 2, 0] AS score

FROM animals

ORDER BY score ~~ASC~~ | [✓ DESC] ;l

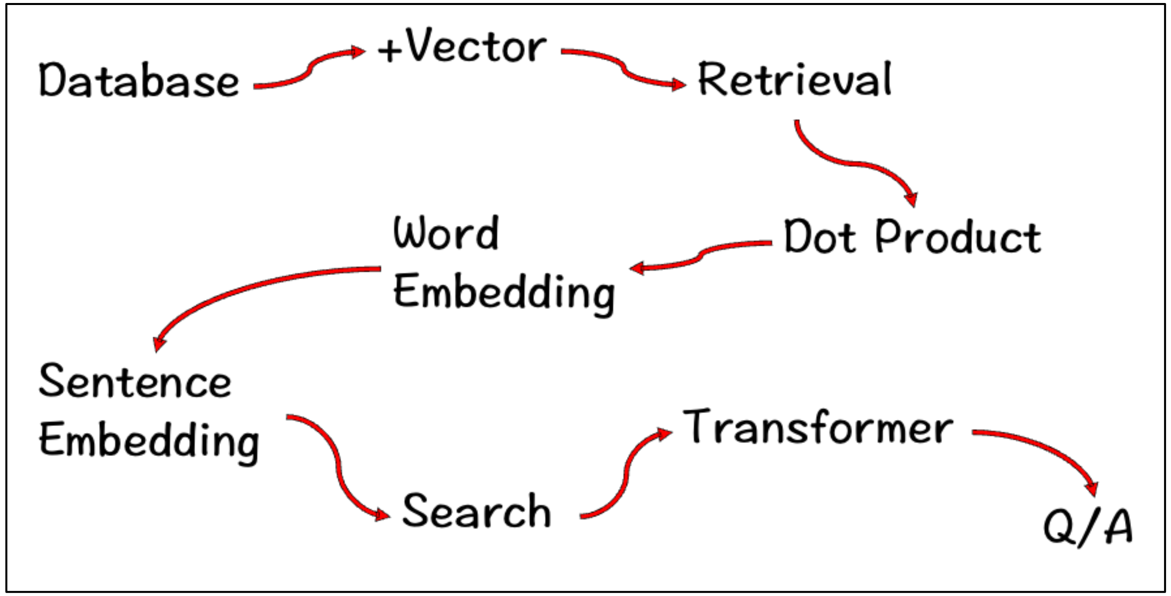
How to retrieve by distance? (Euclidean) L2

- → minus

SELECT name, emb<~~*~~>[1, 2, 0] AS score

FROM animals

ORDER BY score D~~ES~~C; → ASC



Database

+Vector

Retrieval

Dot Product

Word Embedding

Sentence Embedding

Search

Transformer

Q/A

Dot Product

Beginner’s Guide to Vector Databases - AI by Hand ✍️

University of Colorado
Boulder

How to compute dot product?

Example:

A	1	2	3	
	*	*	*	
B	2	2	0	
	=	=	=	Σ
	2	4	0	6
				Result

dog	2	1	0	
	*	*	*	
cat	1	2	0	
	=	=	=	Σ
	2	2	0	4
				Result

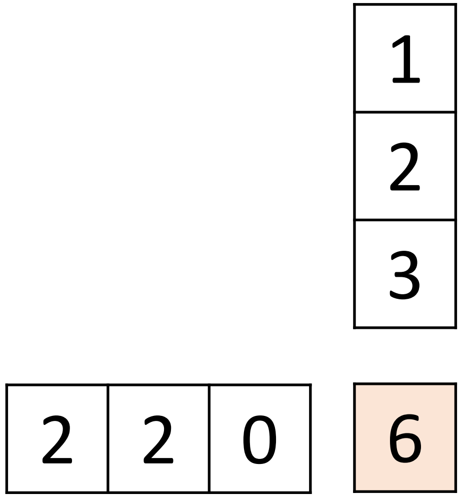
↓

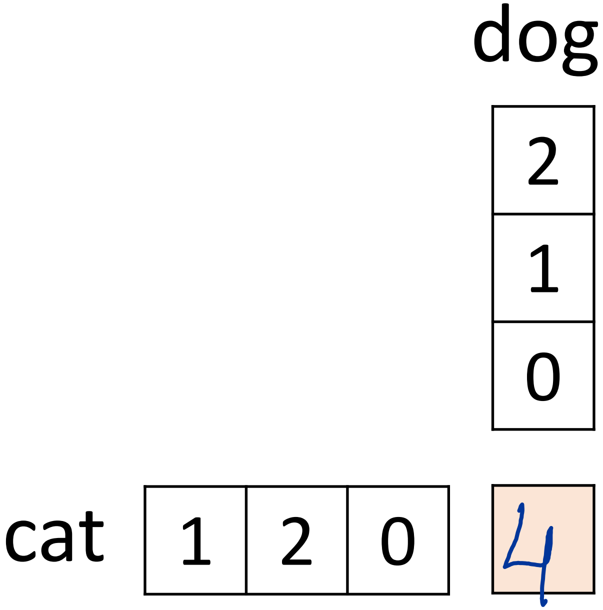
similarity

How to compute dot product using matrix multiplication?

→ scalable

Example:



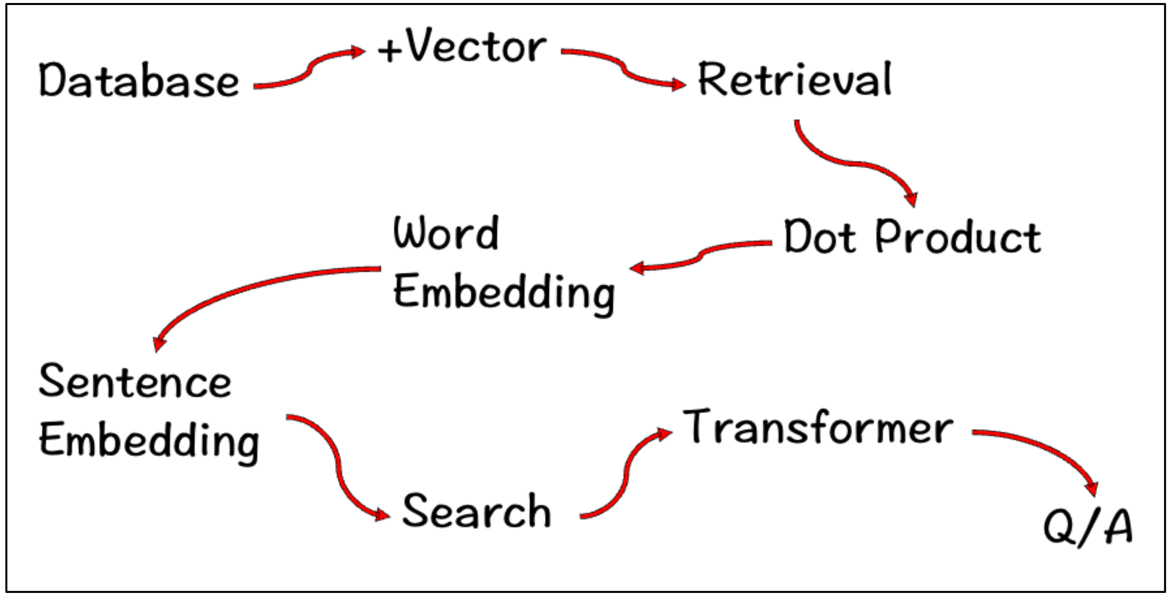


How to compute dot product with multiple vectors?

Example:

$$\begin{array}{ccccc} & & & 1 & 1 \\ & & & 2 & 1 \\ & & & 3 & 1 \\ 2 & 2 & 0 & 6 & 4 \end{array}$$

$$\begin{array}{cccccc} & & & & \text{dog} & \text{bat} \\ & & & & 2 & 0 \\ & & & & 1 & 1 \\ & & & & 0 & 2 \\ \text{cat} & 1 & 2 & 0 & 4 & 2 \end{array}$$

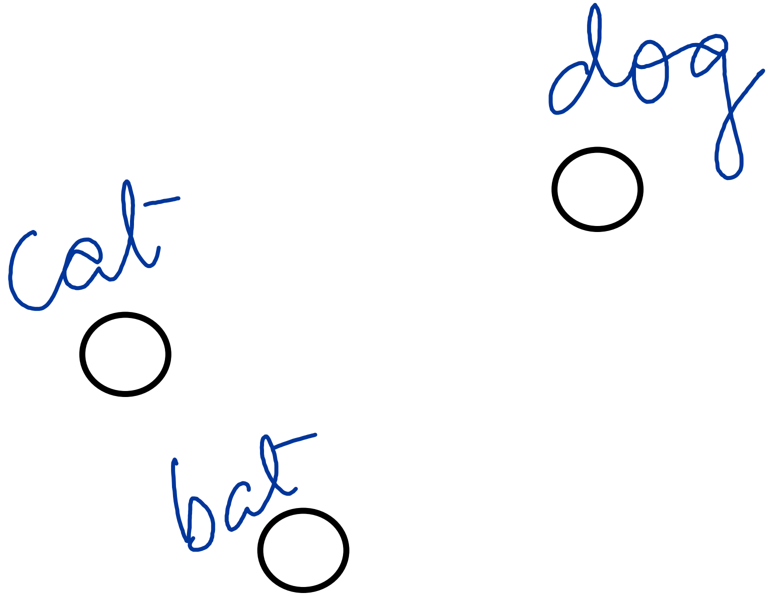


Word Embedding

Beginner's Guide to Vector Databases - AI by Hand ✍️

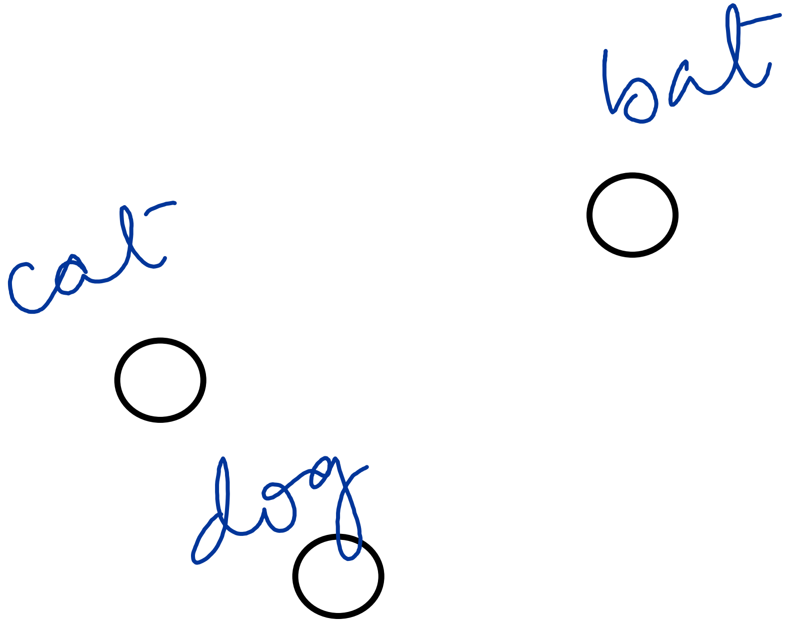
University of Colorado
Boulder

Where are dog, cat and bat in the "name" space?



Where are dog, cat and bat in the “na~~m~~e” space?

semantic
space



Which embedding is better?

Embedding 1

dog	cat	bat
2	1	0
1	2	1
0	0	2

Embedding 2

dog	cat	bat
2	0	1
1	1	0
0	2	2

Which embedding is better?

Desired dot product similarity

dog	cat	bat
2	1	0
1	2	1
0	0	2

dog	2	1	0
cat	1	2	0
bat	0	1	2

	H	L
H		L
L	L	

we want this

H ⇒ high

L ⇒ low

Embedding 1

dog	cat	bat
2	1	0
1	2	1
0	0	2

dog	2	1	0
cat	1	2	0
bat	0	1	2

	4	1
4		2
1	2	

3 x 3 matrix

Embedding 2

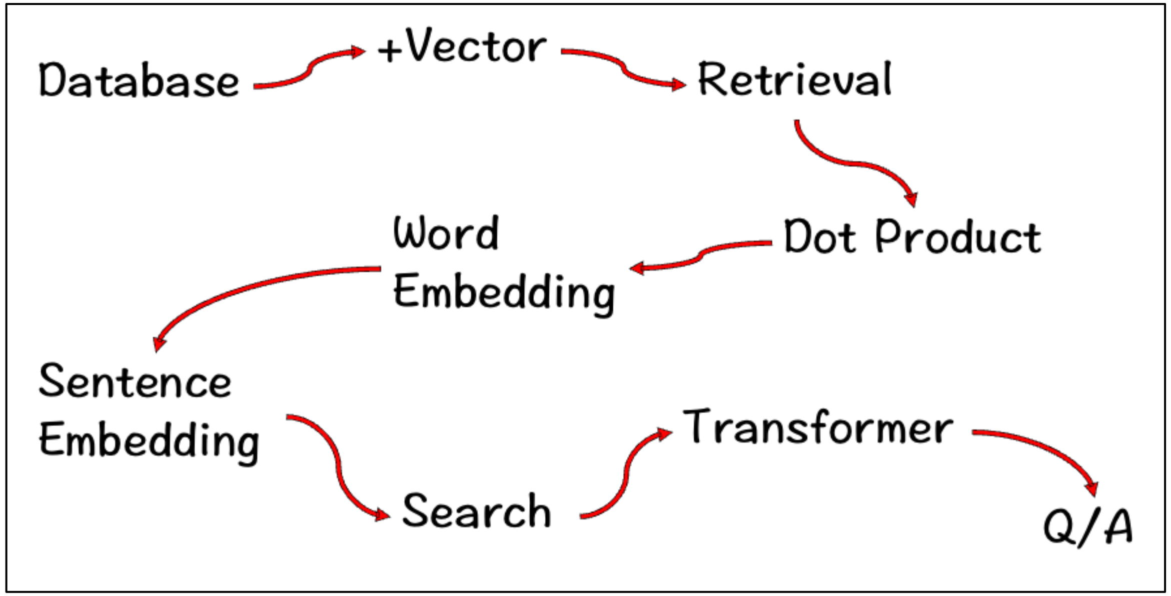
dog	cat	bat
2	0	1
1	1	0
0	2	2

dog	2	1	0
cat	0	1	2
bat	1	0	2

	1	2
1		4
2	4	

⇓

Yellow is better



Sentence Embedding

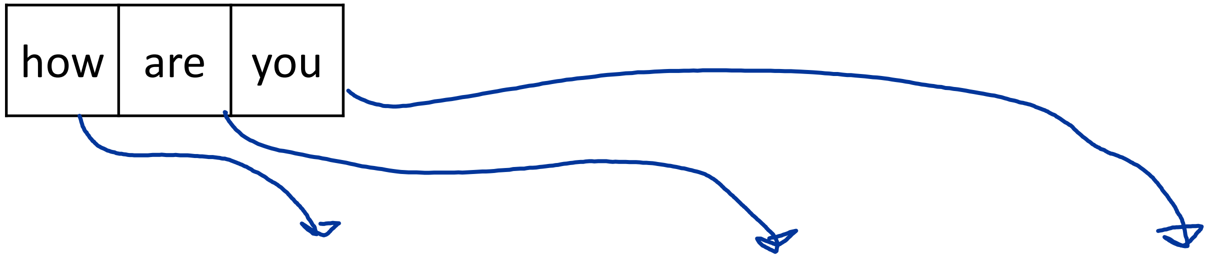
Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

How to embed sentences?

id	comment	user	emb
1	How are you?	John	?
2	Who are you?	Mary	?

“How are you” → word embedding vectors



a	an	the	how	why	who	what	are	is	am	be	was	you	we	I	they	she	he	she	me	him	her
0	-1	0	1	0	1	0	0	-1	1	0	0	0	3	1	0	-1	0	0	0	-1	0
2	0	2	0	0	0	-1	1	0	0	0	2	1	0	2	0	2	0	0	2	0	0
-1	0	-1	1	2	0	0	1	0	1	-1	0	0	-1	0	3	0	0	-1	0	2	-1
0	1	0	0	1	0	1	0	1	0	1	-2	0	0	0	1	0	1	0	1	0	1

1	0	0
0	1	1
1	1	0
0	0	0

Word vectors → Sentence vector

Method 1: Concatenate

how	are	you
1	0	0
0	1	1
1	1	0
0	0	0

1
0
1
0
0
1
1
0
0
1
0
0

→ This might not be scalable

Word vectors → Sentence vector

Method 2: Average

$$\begin{array}{|c|c|c|} \hline \text{how} & \text{are} & \text{you} \\ \hline 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ \hline \end{array} \qquad \begin{array}{|c|} \hline 1/3 \\ 2/3 \\ 2/3 \\ 0/3 \\ \hline \end{array}$$

Vector(4)

id	comment	user	emb
1	How are you?	John	[1/3, 2/3, 2/3, 0/3]
2	Who are you?	Mary	

“Who are you” → word embedding vectors

who	are	you

a	an	the	how	why	who	what	are	is	am	be	was	you	we	I	they	she	he	she	me	him	her
0	-1	0	1	0	1	0	0	-1	1	0	0	0	3	1	0	-1	0	0	0	-1	0
2	0	2	0	0	0	-1	1	0	0	0	2	1	0	2	0	2	0	0	2	0	0
-1	0	-1	1	2	0	0	1	0	1	-1	0	0	-1	0	3	0	0	-1	0	2	-1
0	1	0	0	1	0	1	0	1	0	1	-2	0	0	0	1	0	1	0	1	0	1

1	0	0
0	1	1
0	1	0
0	0	0

Word vectors → Sentence vector

Method 2: Average

$$\begin{array}{ccc} \text{who} & \text{are} & \text{you} \\ \hline 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \qquad \begin{array}{c} \\ \hline 1/3 \\ 2/3 \\ 1/3 \\ 0/3 \end{array}$$

id	comment	user	emb
1	How are you?	John	[1/3, 2/3, 2/3, 0]
2	Who are you?	Mary	[1/3, 2/3, 1/3, 0]

How to query by SQL?

dot product/similarity

SELECT comment, emb<*>[2/3, 1/3, 3/3, 4/3] AS score

FROM posts

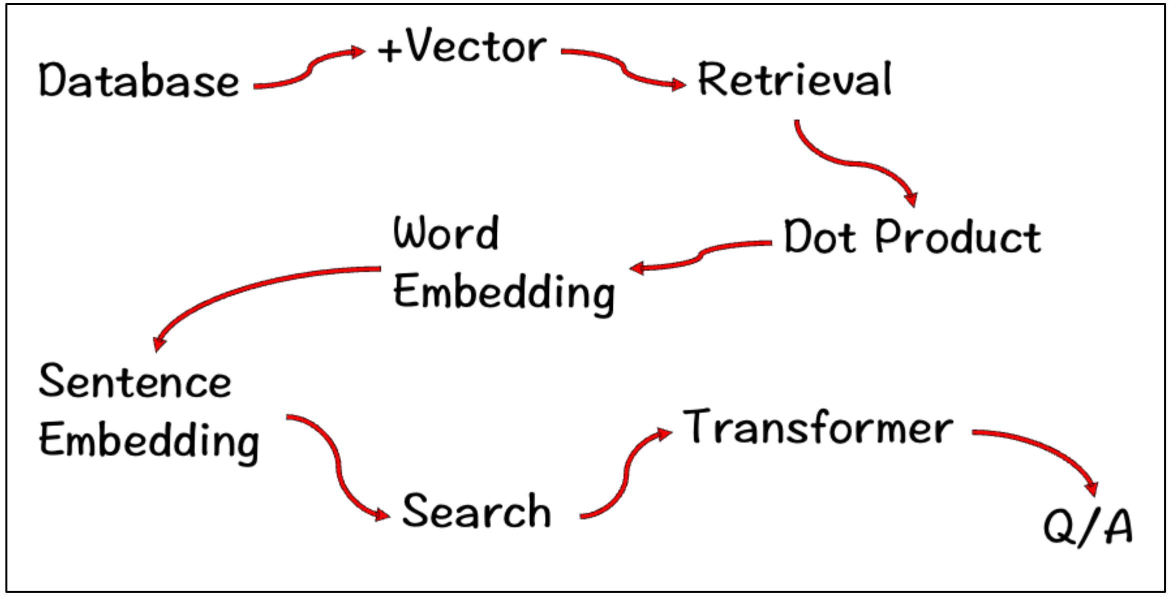
ORDER BY score ~~ASC~~ | DESC ✓ ;

How to query using a high-level API?

```
query = Query(post_index)
      .find(post)
      .similarity(relevance_space.text, Param("comment"))

app.query(query, comment = "who are you?" )
```

Source: Superlinked.com



Database → +Vector → Retrieval → Dot Product → Word Embedding → Sentence Embedding → Search → Transformer → Q/A

Search

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado
Boulder

K-Nearest Neighbor, K=3, Dot-Product

Database

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
emb																				

Query				

9	-8	9	9	0	3	1	-6	0	11	3	13	-2	6	15	-9	7	6	-5	8

{ [✓ max] | min }										3		2			1					

K-Nearest Neighbor, K=3, L2 → Euclidean Dist (-)

Database

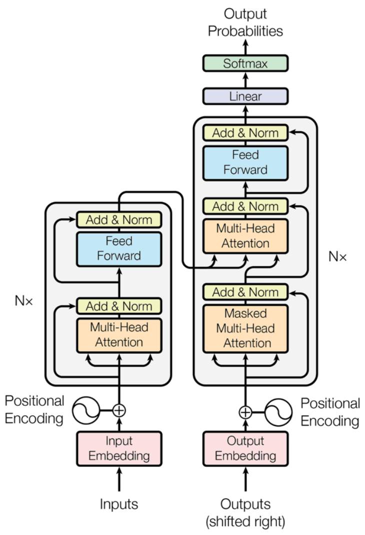
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
emb																				

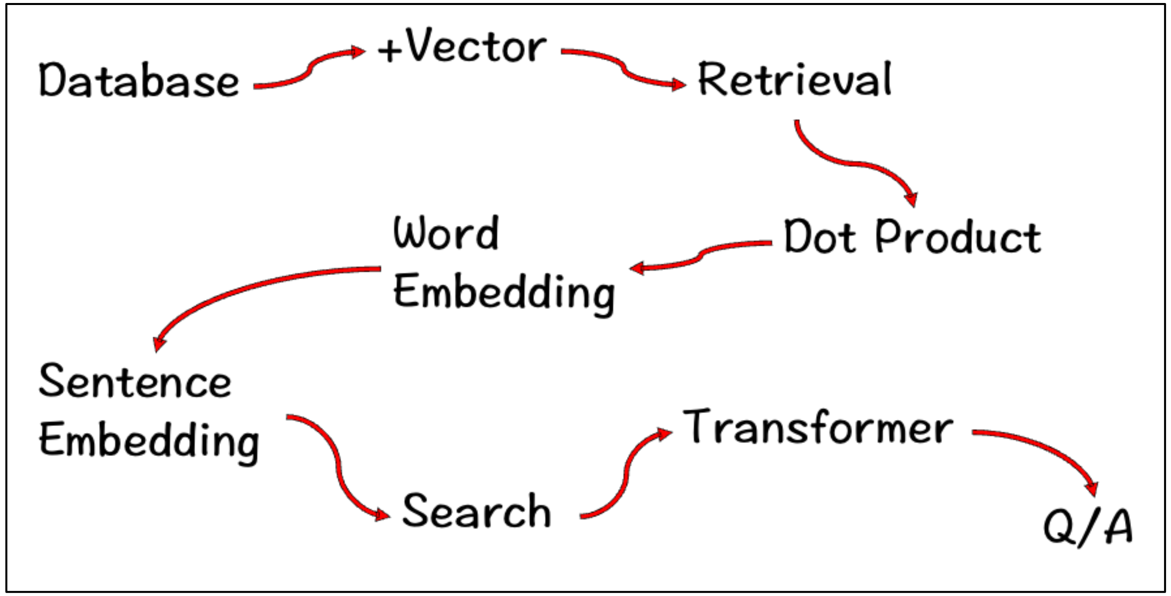
Query				

	6	8	9	9	1 ✓	10	0 ✓	9	12	15	2	13	12	6	15	9	7	6	5	8
{ max [✓ min] }					2		1				3									

* 20m records ⇒ this process will be slow

* ANN ⇒ retrieval is faster





Transformer

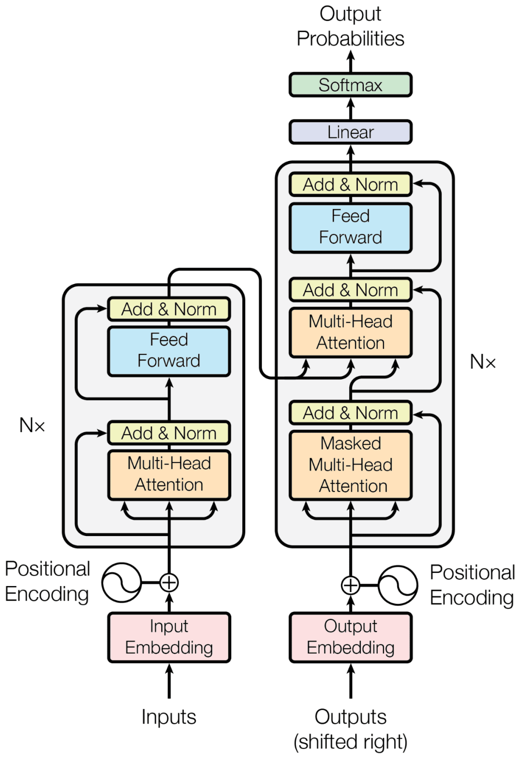
Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder

How to use a Transformer to get a sentence embedding vector?

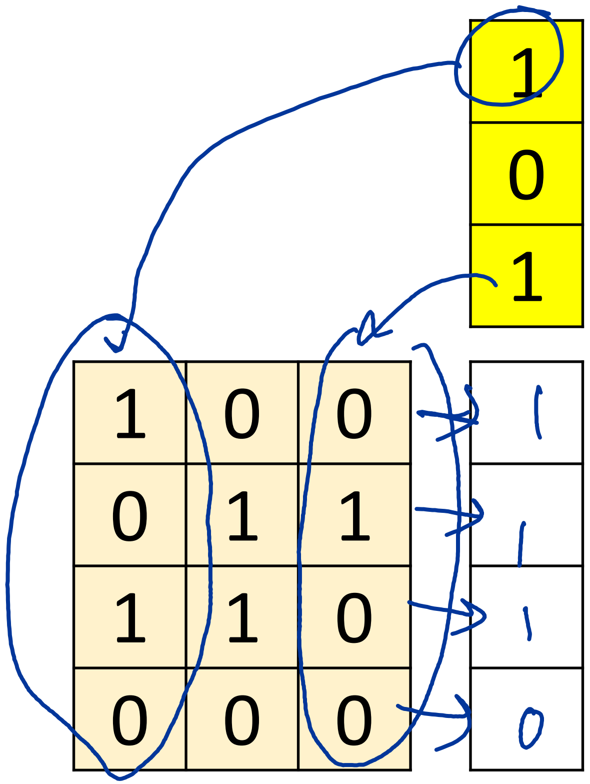
Word Embedding Vectors

1	0	0
0	1	1
1	1	0
0	0	0



Sentence Embedding Vector

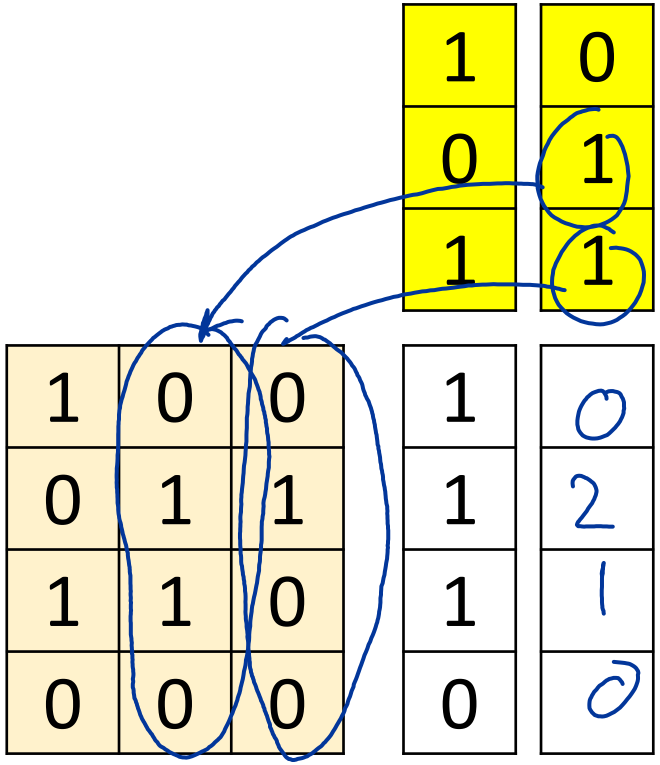
How to combine across positions?



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

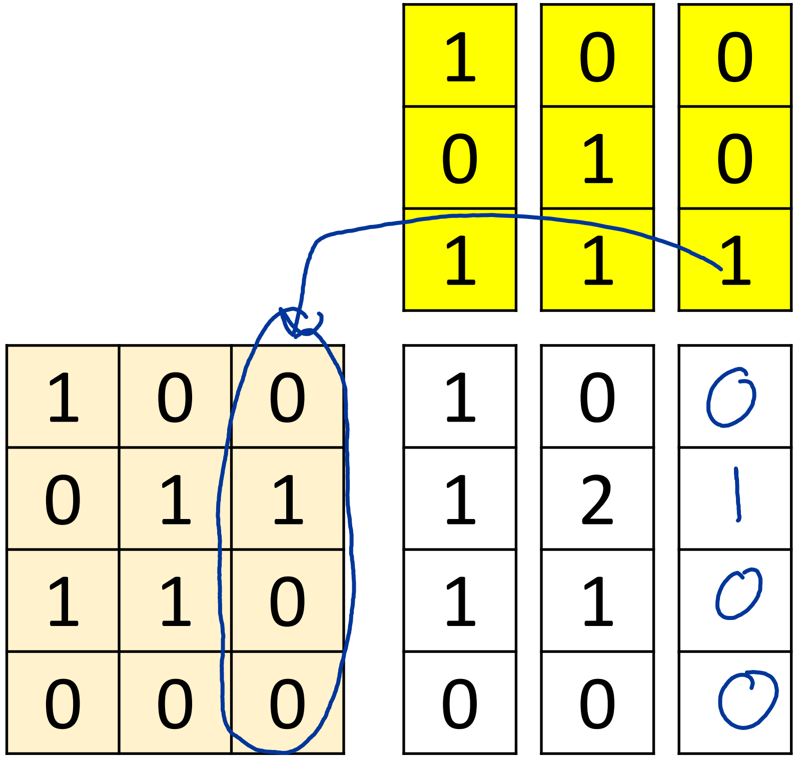
$$\begin{bmatrix}1&0&0\\0&1&1\\1&1&0\\0&0&0\end{bmatrix} \quad \begin{bmatrix}1\\1\\1\\0\end{bmatrix}$$

How to combine across positions?



$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$$

How to combine across positions?

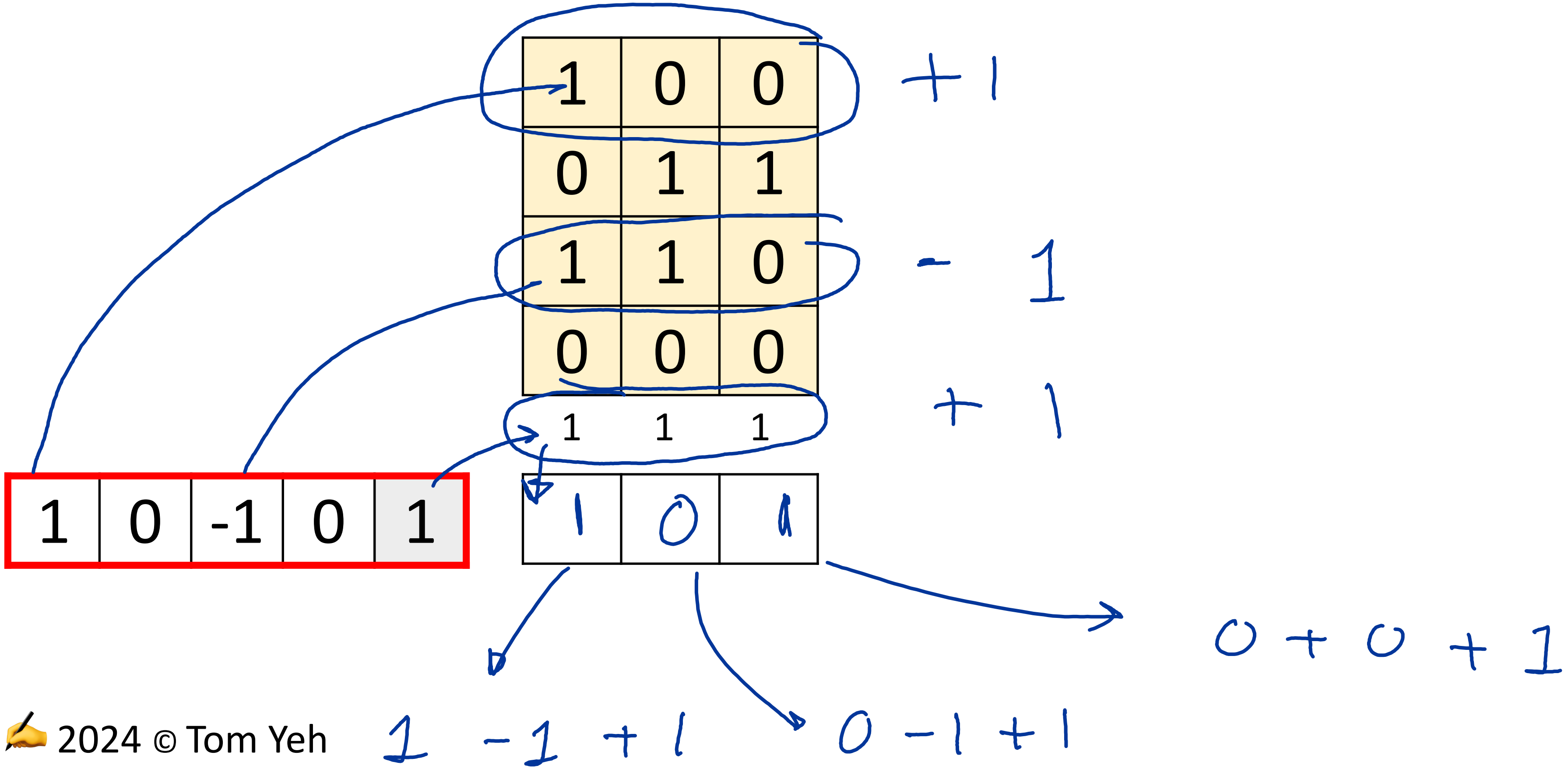


1	0	0
0	1	0
1	1	1

1	0	0
0	1	1
1	1	0
0	0	0

1	0	0
1	2	1
1	1	0
0	0	0

How to combine across features?

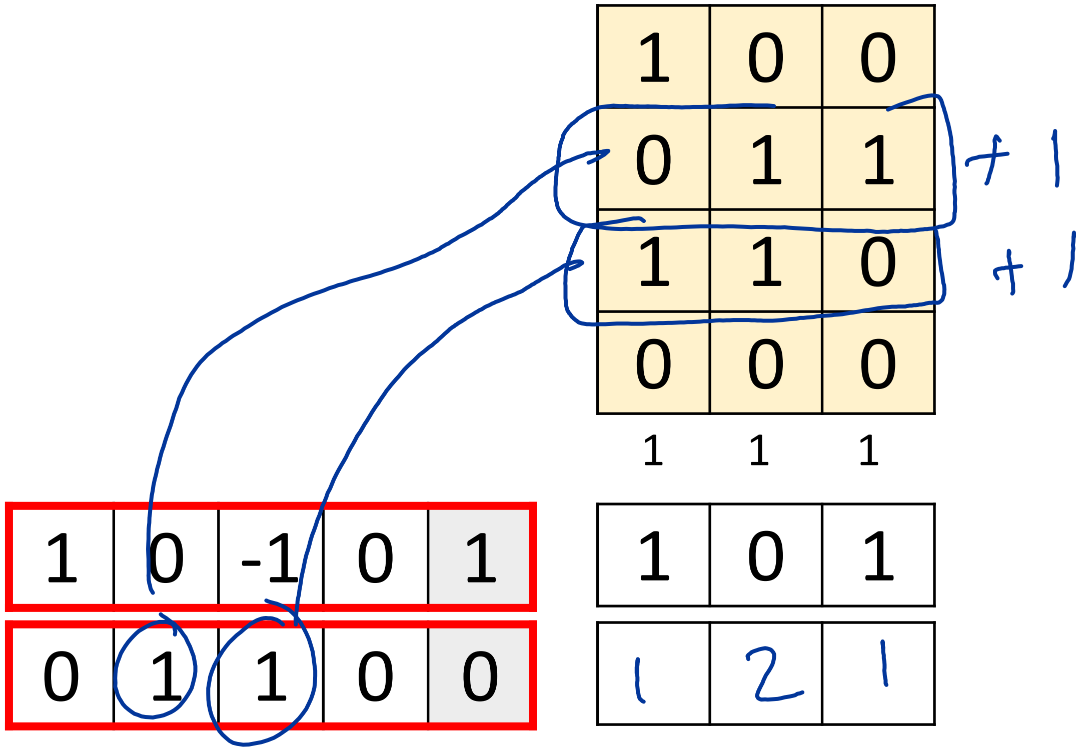


1	0	-1	0	1

1	0	0
0	1	1
1	1	0
0	0	0
1	1	1

1	0	1

How to combine across features?

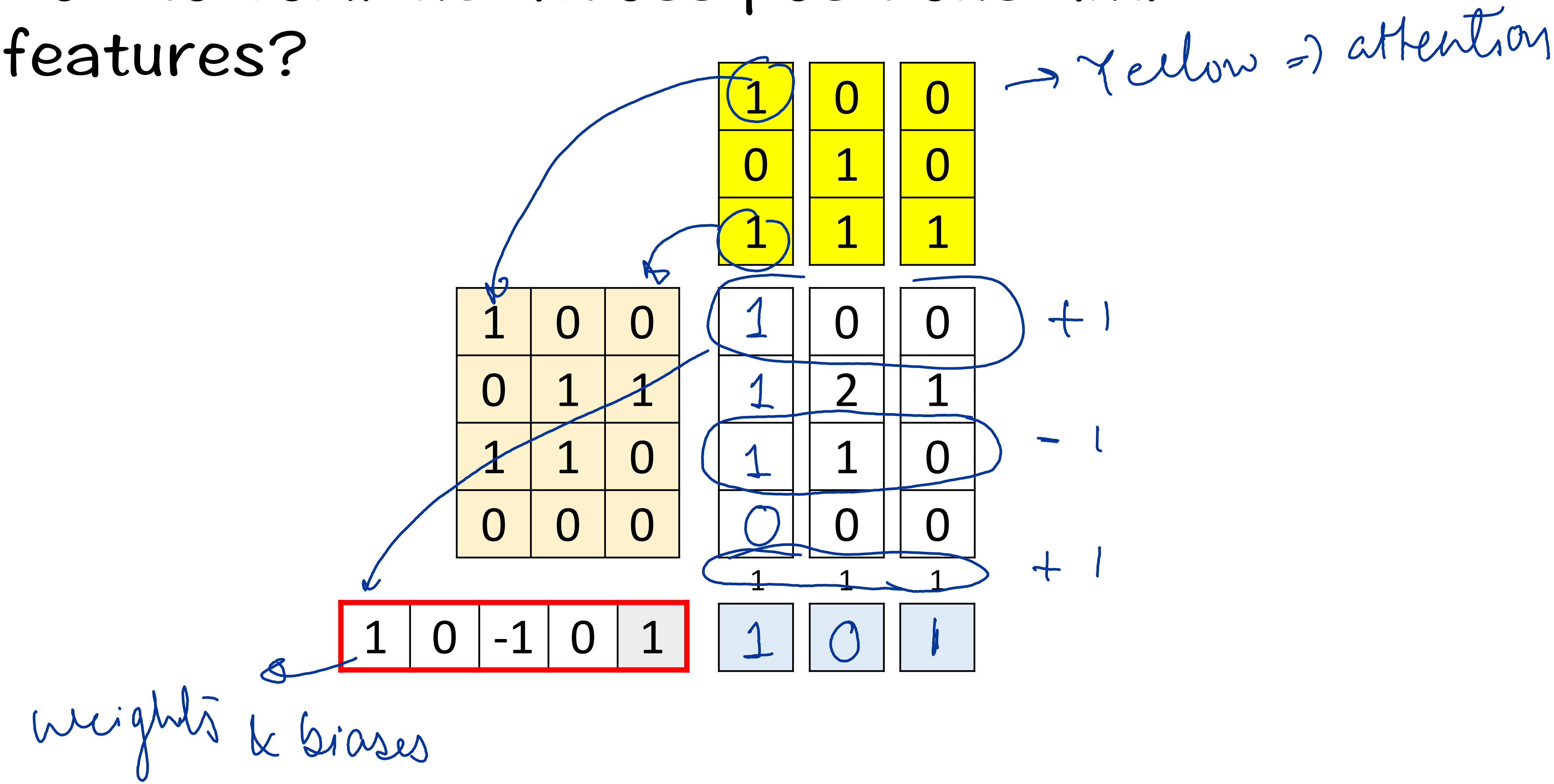


1	0	0
0	1	1
1	1	0
0	0	0
1	1	1

1	0	-1	0	1
0	1	1	0	0

1	0	1
1	2	1

How to combine across positions and features?



→ Yellow ⇒ attention

1	0	0
0	1	0
1	1	1

1	0	0
0	1	1
1	1	0
0	0	0

1	0	0	+1
1	2	1	
1	1	0	-1
0	0	0	
1	1	1	+1

1	0	-1	0	1

1	0	1

weights & biases

How to use a Transformer to get a sentence embedding vector?

Word Embedding Vectors

1	0	0
0	1	0
1	1	1

1	0	0
0	1	1
1	1	0
0	0	0

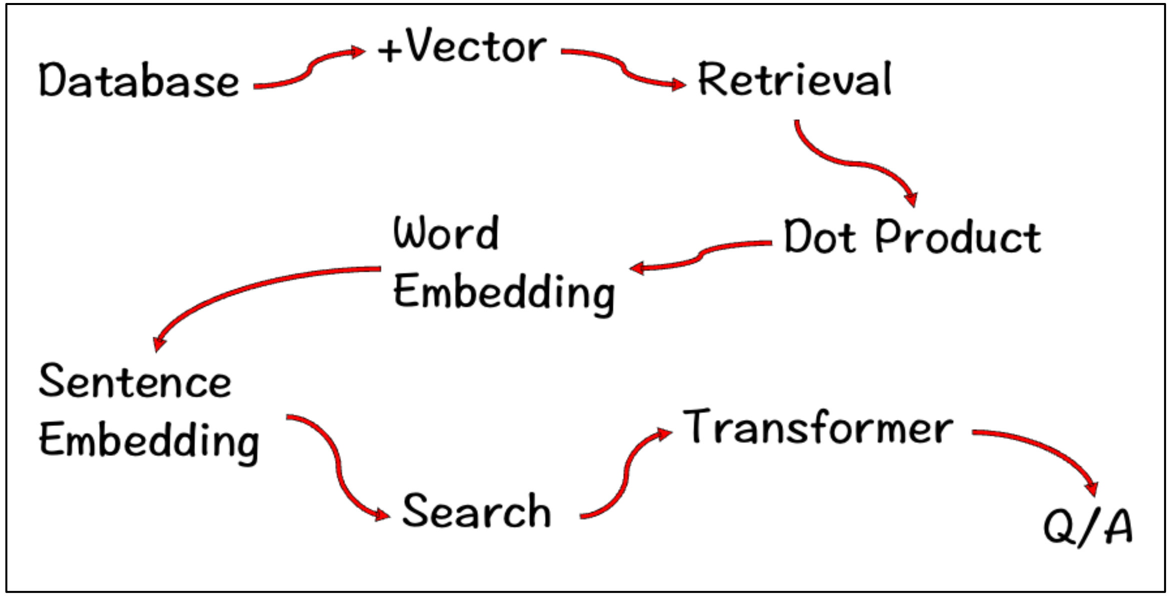
1	0	0
1	2	1
1	1	0
0	0	0
1	1	1

1	0	-1	0	1
0	1	1	0	0
0	0	0	1	1
0	0	1	1	0

1	0	1
2	3	1
1	1	1
1	1	0

Sentence Embedding Vector

2/3
6/3
3/3
2/3



Q/A

Beginner's Guide to Vector Databases - AI by Hand ✍️

University of Colorado Boulder