Vector Databases

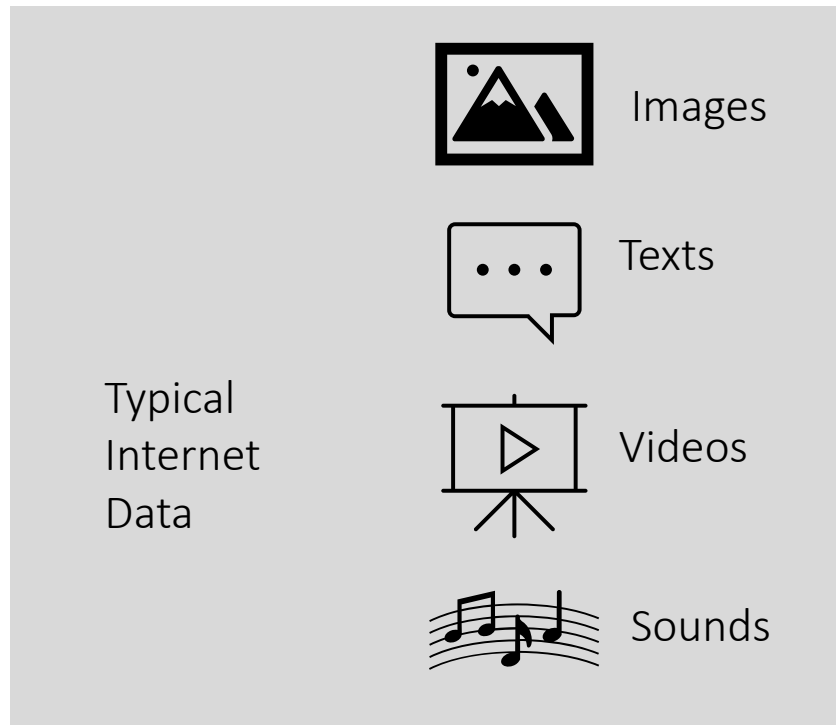# Vector Databases

- Special type of database
- Store, manage, and query data which is represented in geometric formats
- Very famous in Deep Learning (NLP)
- Allow
  - similarity search
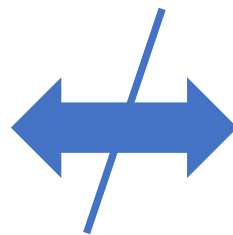 - Clustering
  - Real-time analytics

A vector database stores high-dimensional data (embeddings) for fast querying and similarity analysis.
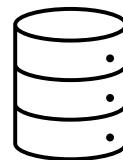
Vector Databases
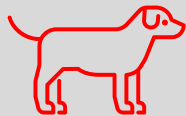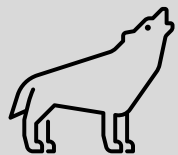
Why are Vector Databases required?

Typical Internet Data

Images

Texts

Videos

Sounds

Typical Databases

SQL

Unstructured Data

Structured Data
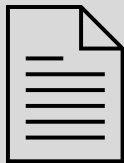
Vector Databases

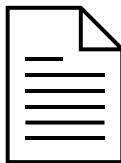Candidate Images

Query Image

Task: Which image is most similar?

Vector Databases

Example: Text Similarity

Candidate Words/Texts

Query Words/Texts
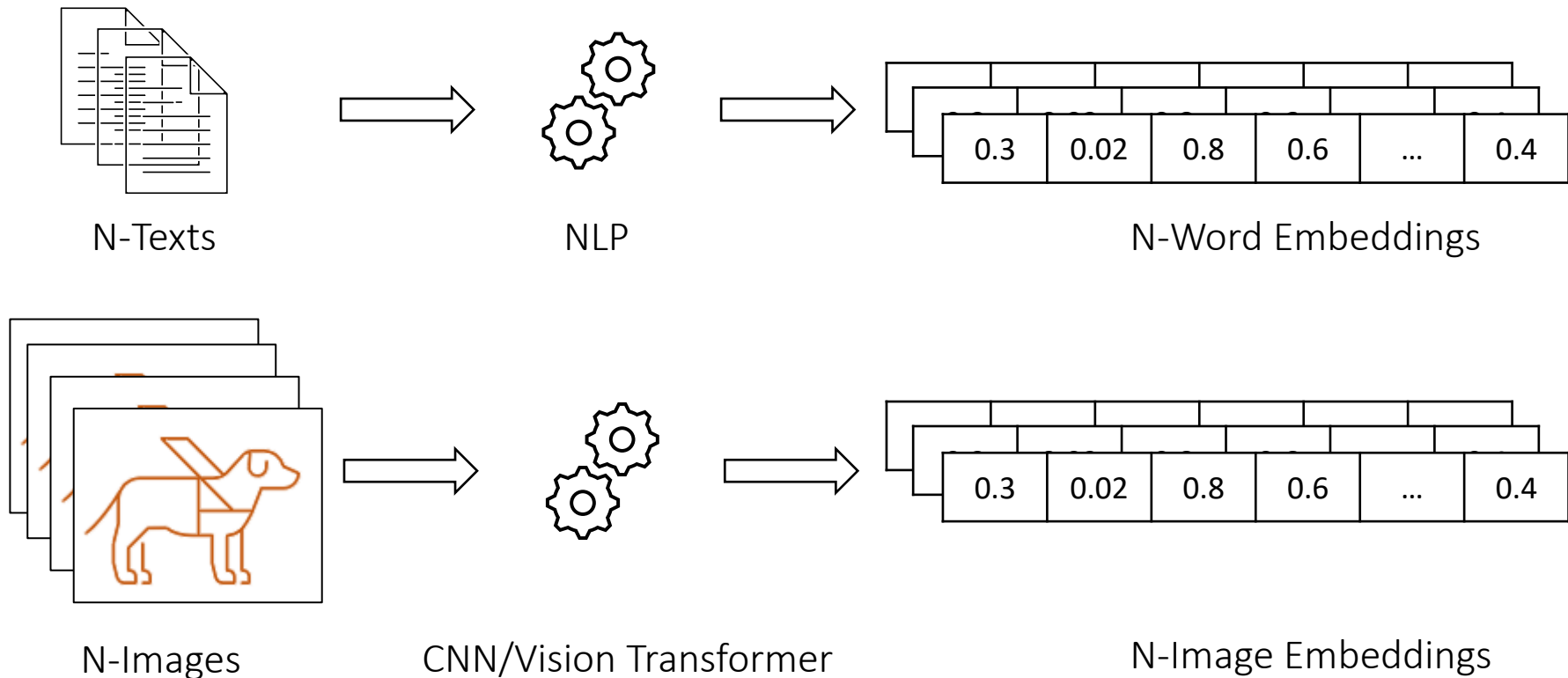
Task: Which text is most similar?

Vector Databases

Embeddings

N-Texts → NLP → N-Word Embeddings

| 0.3 | 0.02 | 0.8 | 0.6 | ... | 0.4 |

N-Images → CNN/Vision Transformer → N-Image Embeddings

| 0.3 | 0.02 | 0.8 | 0.6 | ... | 0.4 |

Vector Databases

How to create image embeddings?

CONV | Batch Norm | ReLU | Max Pool | ... | Avg Pool | Flatten | Fully Connected | Output
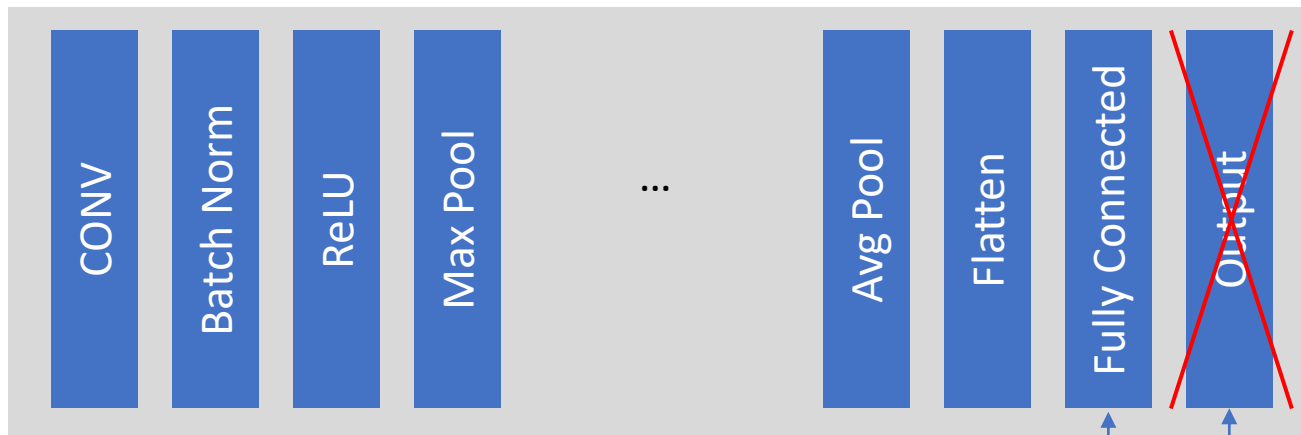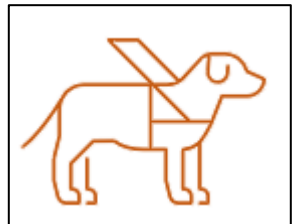
Classes 0-999

Last hidden state

Vector Databases

Find most similar data

GloVe Word Embeddings and Categories
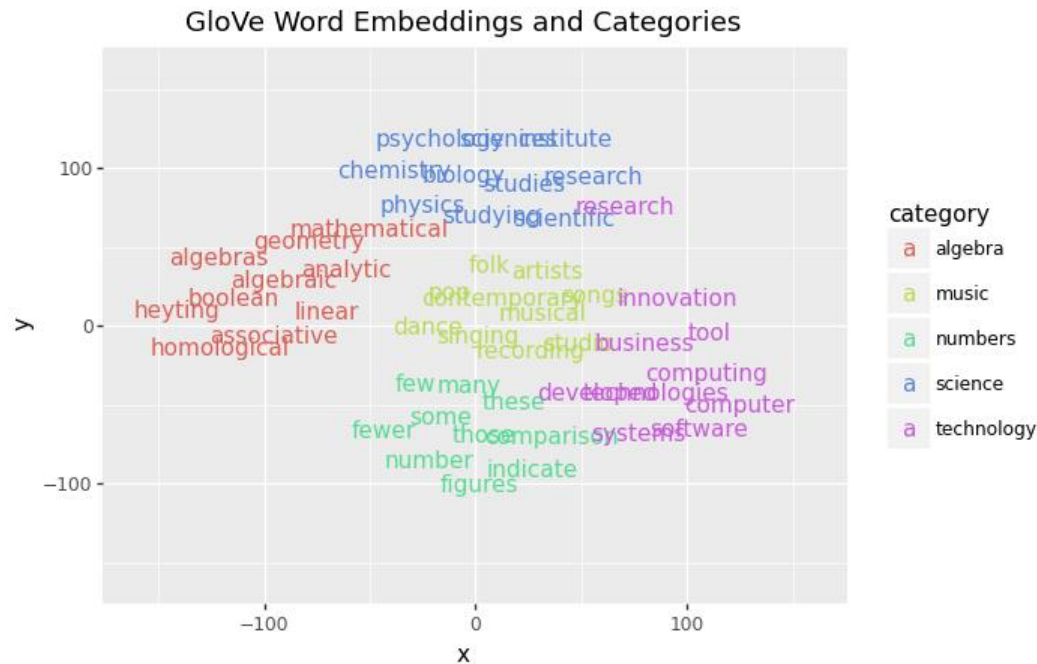
$$dist = \sqrt{(x_1 - y_1)^2 + (x_n - y_n)^2}$$

For an embedding vector of 768 embeddings, there are 768 distance terms

Example: word embeddings reduced to 2 dimensions

Vector Databases

Imagename	Embedding					
dog1	0.3	0.02	0.8	0.6	...	0.4
dog2	0.1	0.52	0.7	0.6	...	0.4
...						
dogN	0.3	0.62	0.9	0.2	...	0.3

Vector Database

dogTest

0.3	0.02	0.8	0.6	...	0.4

$$dist = \sqrt{\sum (x_i - y_i)^2}$$

For small data, go with np.array().

For Big Data this gets slow.

Vector Databases

Imagename	Embedding						Index
dog1	0.3	0.02	0.8	0.6	...	0.4	
dog2	0.1	0.52	0.7	0.6	...	0.4	
...							
dogN	0.3	0.62	0.9	0.2	...	0.3	

Vector Database

Index speeds up search.

Vector Databases

pinecone.io

trychroma.com

redis.com