Vectors, Embeddings, Similarity Search And Pinecone VectorDB

...

Vectors

Vector can be considered as list of numbers.

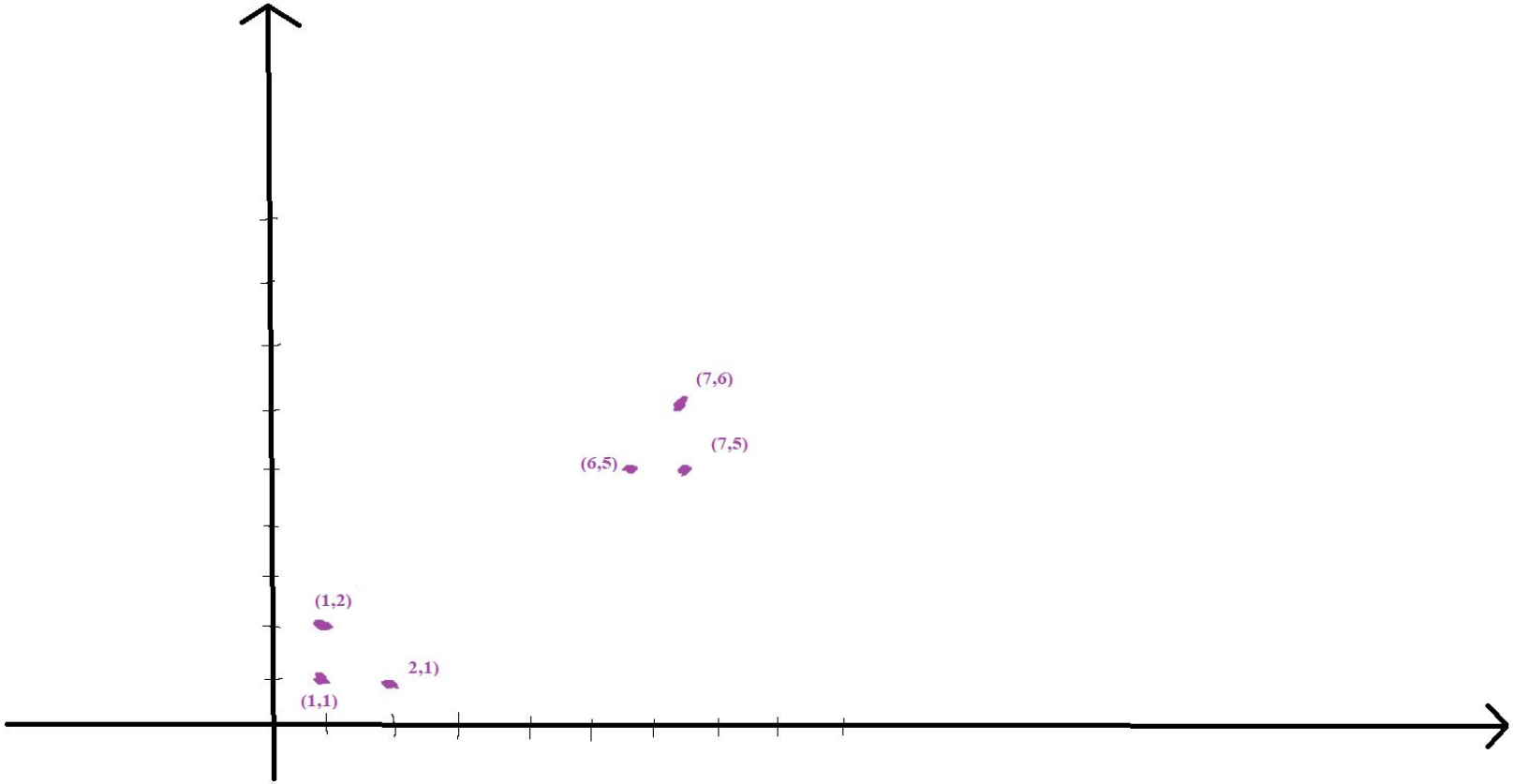
Example:- [1,2,3] and [1,5] and [1,5,7,6,4,6,4]

Dimension of vector => Number of elements in vector is dimension of vector.

Example:-

[1,2,3] is vector in 3 dimensions. Similarly [1,2] is vector in 2 dimensions.

Vectors In 2D And Geometric Similarity



Other Similarity Metrics

Geometric Similarity: In a geometric sense, two vectors are similar if they have similar directions and magnitudes. For example, if you have two velocity vectors representing the speed and direction of two moving objects, and these vectors are close in both direction and magnitude, they could be considered similar.

Cosine Similarity: In the context of natural language processing and information retrieval, vectors can represent documents or words in a high-dimensional space. Cosine similarity is often used to measure the cosine of the angle between two vectors. If the angle is small (cosine value close to 1), the vectors are similar in the sense that they point in roughly the same direction. This is used, for instance, in document similarity or recommendation systems.

Euclidean Distance: The distance between vectors in a multi-dimensional space can also be used to determine their similarity. Smaller distances usually indicate higher similarity. This is often used in clustering and classification algorithms.

Embeddings

Embeddings

An embedding is a vector (list) of floating point numbers.

The distance between two vectors measures their relatedness.

Small distances suggest high relatedness and large distances suggest low relatedness.

Embeddings And Similarity Search Links

https://huggingface.co/blog/getting-started-with-embeddings

https://platform.openai.com/docs/guides/embeddings/what-are-embeddings

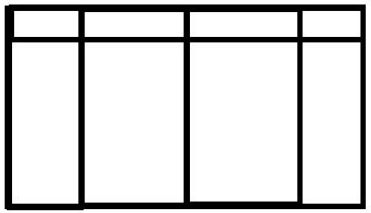
http://vectors.nlpl.eu/explore/embeddings/en/associates/

(dog,cat,mouse,cow,wolf,tiger,lion,cheetah,leopard,lizard,spiderman)

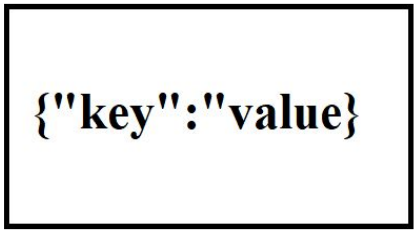
VectorDB

Database to efficiently store and do operation on vectors.

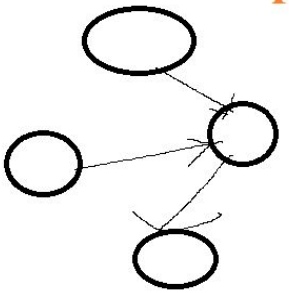
MySQL



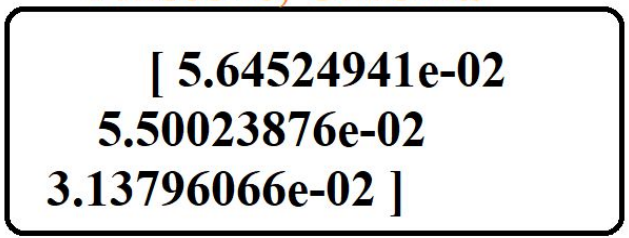
MondoDB



AWS Neptune

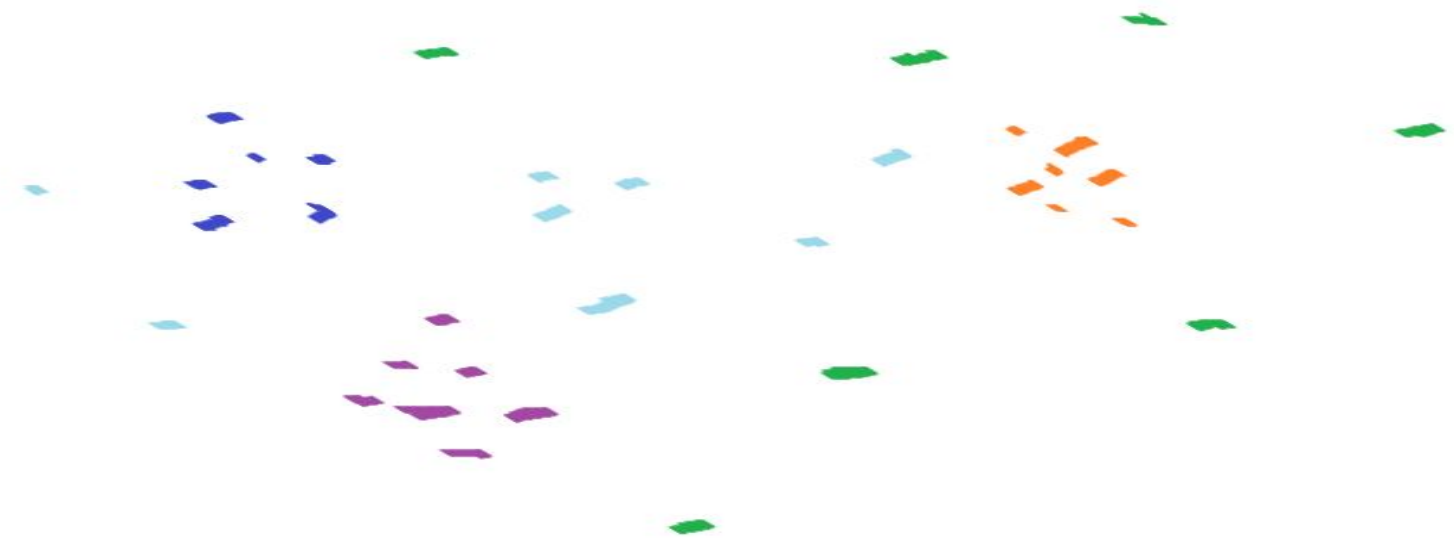


Pinecone, Chroma



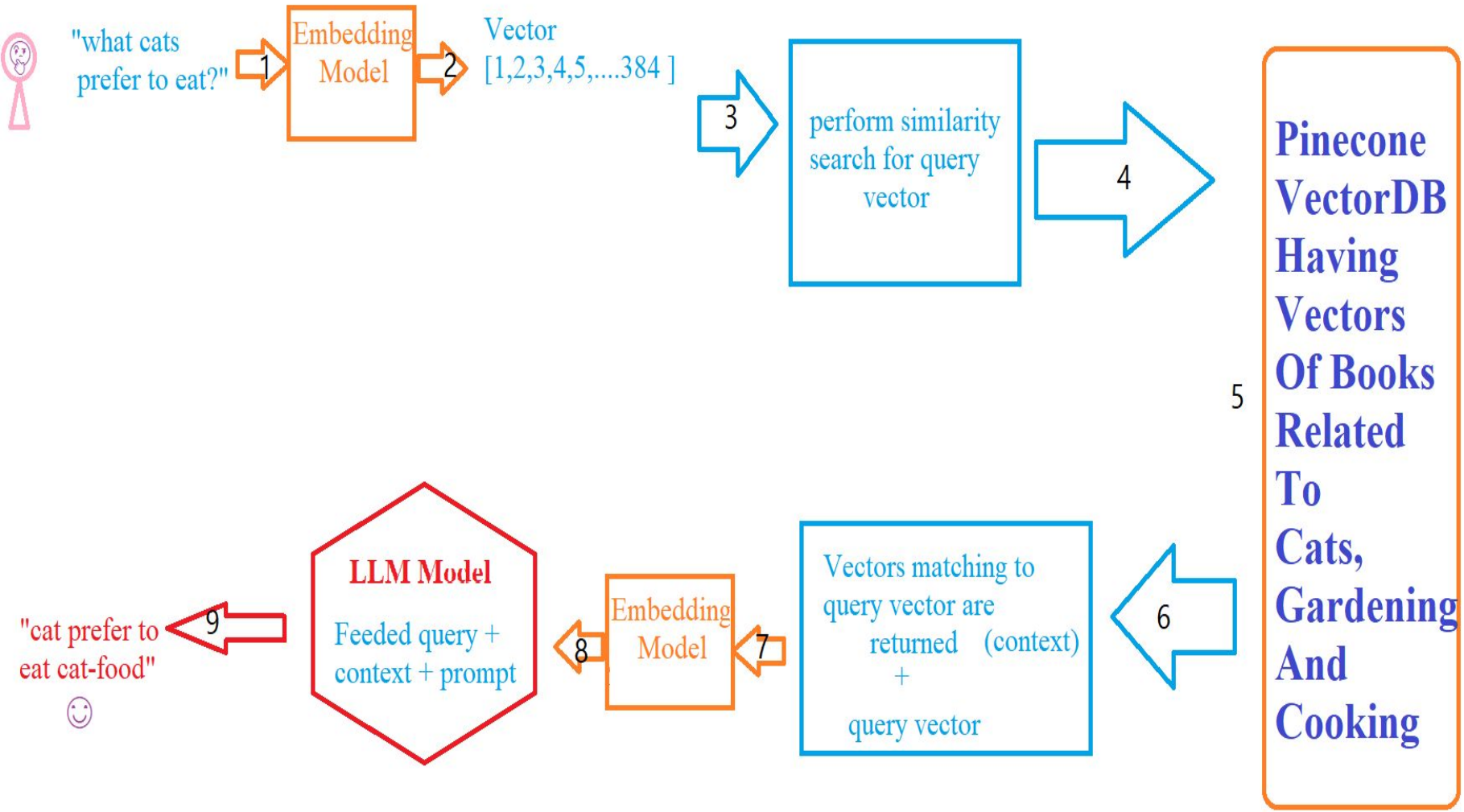
Similarity Search

Embedding model places words having same semantic meaning closer together in vector space. So similar vectors have similar semantic meaning.



Use Case: Retrieval Augmented Generation (RAG)

https://docs.aws.amazon.com/sagemaker/latest/dg/jumpstart-foundation-models-customize-rag.html


"what cats prefer to eat?"
1
Embedding Model
2
Vector [1,2,3,4,5,....384]
3
perform similarity search for query vector
4
Pinecone VectorDB Having Vectors Of Books Related To Cats, Gardening And Cooking
5
6
Vectors matching to query vector are returned (context) + query vector
7
Embedding Model
8
LLM Model
Feeded query + context + prompt
9
"cat prefer to eat cat-food"

Let’s Play With Pinecone

…

Initialize PineCone

```
import pinecone


pinecone.init(
        api_key= “”
        environment='us-west1-gcp-free'
)
index = pinecone.Index(")
```

Pinecone: Upsert And Query

```
index.upsert(vectors=zip(["1", "2", "3"], new))
```

```
index.query(queries=embeddings, top_k=3, include_values=True)
```

Thank You 😀

...