CODE BASICS

Vector Database

TAKEAWAYS

codebasics.io

Chromadb: Introduction and Installation

1. Vector databases store data as vectors (embeddings), enabling fast and efficient similarity searches across high-dimensional spaces.
2. With Gen AI revolution, many vector databases have come up. Here are the few popular ones: Chromadb (open source), Milvus (opensource), Pinecone, Qdrant

Euclidean and Cosine Distance

1. Euclidean distance measures the straight-line distance between two points in a vector space, reflecting absolute similarity.
2. Cosine distance evaluates the angle between two vectors, focusing on their orientation rather than magnitude.
3. Euclidean distance is sensitive to vector magnitude, while cosine distance emphasizes direction, making it ideal for text and document similarity.
4. Choosing between Euclidean and cosine distance depends on the task – Euclidean for dense embeddings and cosine for sparse or high-dimensional data.