Lecture 19: What is Cosine Similarity

Definition:

Cosine similarity is a mathematical method to measure how similar two vectors are by calculating the cosine of the angle between them. It is widely used in vector databases and embeddings to compare semantic similarity.

Concept:

- Embeddings represent words/text as vectors (multi-dimensional values).
- Similar vectors (e.g., happy and joy) are close to each other.
- Opposite vectors are placed far apart.
- Cosine similarity finds how "close" or "far" two vectors are based on their angle.

Formula for Cosine Similarity:

$$\text{cosine similarity} = \frac{A \cdot B}{||A|| \cdot ||B||}$$

- A · B → Dot product of vectors A and B.
- ||A||, ||B|| → Magnitudes of vectors.
- **Range**:
  - 1 → Identical direction (high similarity)
  - 0 → Orthogonal (no similarity)
  - -1 → Opposite direction

How to implement it in code:

- In Python (NumPy) and libraries, cosine similarity is built-in.
- But, in Java:
  - No direct method by default.
  - Can write a reusable function using the dot product and magnitude formula.
- For multi-dimensional vectors, use summation with a loop to calculate dot product and magnitudes.