Day 3: Vector Operations — Dot Product, Norms, Cosine Similarity & Projections

Goal. Understand core vector operations used everywhere in AI: dot product, norms, cosine similarity, and projections. These power similarity search, recommendation, embeddings, attention, and more.

1. Vectors & Notation

Let $a, b \in \mathbb{R}^n$ be column vectors. The i-th entry of a is a_i. The Euclidean (L2) norm is

$$\|a\| = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

2. Dot Product: Algebraic & Geometric Views

Algebraic:

$$a \cdot b = \sum_{i=1}^{n} a_i b_i.$$

Geometric: If θ is the angle between a and b,

$$a \cdot b = \|a\| \, \|b\| \cos \theta.$$

Thus, the dot product measures *alignment*. Large positive values $\Rightarrow$ small angle (similar direction). Zero $\Rightarrow$ orthogonal (no linear alignment). Large negative $\Rightarrow$ opposite direction.

3. Norms (Magnitude)

The L2 norm measures vector length. Useful normalizations:

$$\hat{a} = \frac{a}{\|a\|}, \qquad \hat{b} = \frac{b}{\|b\|}.$$

Normalization isolates *direction* by removing scale.

4. Cosine Similarity

Definition.

$$\cos\_\mathrm{sim}(a, b) = \frac{a \cdot b}{\|a\|\,\|b\|} = \cos\theta.$$

It depends only on direction, not magnitude.

Proposition 1 (Bounds via Cauchy–Schwarz)**.** *For all $a, b \in \mathbb{R}^n$, $-1 \le \cos\_\mathrm{sim}(a, b) \le 1$.*

Interpretation.

- 1: identical direction (maximally similar).
- 0: orthogonal (no linear alignment).
- -1: opposite directions.

5. Orthogonality & Independence (Caution)

If $a \cdot b = 0$, vectors are *orthogonal* (uncorrelated in a linear sense). This is *not* the same as statistical independence, but orthogonality is a useful proxy for "no linear relation" in feature spaces.

6. Projections

Projection of a onto direction $b \neq 0$:

$$\mathrm{proj}_b(a) = \frac{a \cdot b}{\|b\|^2}\, b.$$

The scalar coefficient $\frac{a \cdot b}{\|b\|}$ is the length of a along the unit direction $\hat{b}$.

7. Worked Numeric Example

Let $a = (1, 2, 3)^\top$, $b = (2, 0, 1)^\top$.

$a \cdot b = 1 \cdot 2 + 2 \cdot 0 + 3 \cdot 1 = 5, \quad \|a\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}, \quad \|b\| = \sqrt{2^2 + 0^2 + 1^2} = \sqrt{5}.$

Cosine similarity:

$$\cos\_\mathrm{sim}(a, b) = \frac{5}{\sqrt{14}\sqrt{5}} \approx \frac{5}{8.3666} \approx 0.598.$$

Projection of a onto b:

$$\mathrm{proj}_b(a) = \frac{5}{5}\, b = b = (2, 0, 1)^\top.$$

Here a has a strong component along b (cosine ≈ 0.6), and its projection equals b because $a \cdot b = \|b\|^2$.

8. Why This Matters in AI

- **Embeddings & Semantic Search:** Compare text/image/audio embeddings with cosine similarity.
- **Recommendation:** User/item vectors; $a \cdot b$ ranks relevance.
- **Attention Mechanisms:** Query/key alignment often reduces to (scaled) dot products.
- **Feature Engineering:** Orthogonality reduces redundancy; projections decompose signals along meaningful directions.

9. Practical Tips

- Use **cosine similarity** when scale should be ignored (e.g., sentence length).
- Use **dot product** when magnitude encodes importance (e.g., TF-IDF with weights).
- Always guard against zero vectors before normalizing.
- In high dimensions, many random vectors are nearly orthogonal; calibrate thresholds accordingly.

10. Mini-Exercises

1. Compute $\text{cos\_sim}\Big((1,0,0),(1,1,0)\Big)$ and interpret.
2. Show that $\|a-b\|^2 = \|a\|^2 + \|b\|^2 - 2\, a \cdot b$ and relate this to similarity vs. distance.
3. For nonzero b, prove $\text{proj}_b(a)$ is the closest point to a on the line $\{tb : t \in \mathbb{R}\}$.

Next

Day 4: Matrix multiplication as linear transformation; transpose, and how backprop uses transposes.