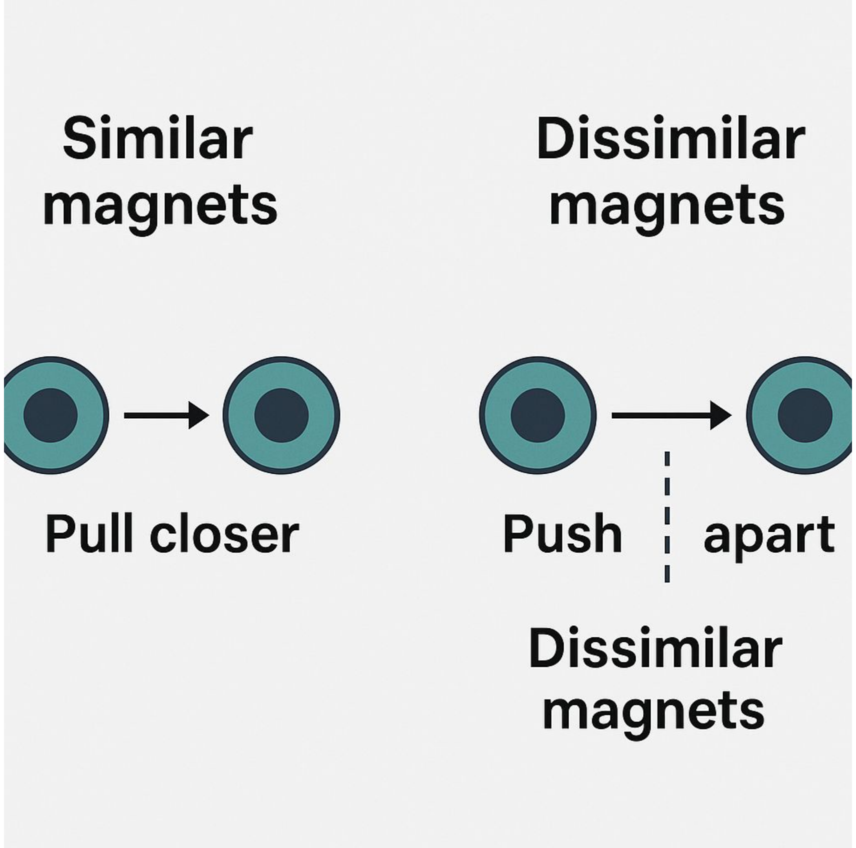
Contrastive Loss Explained



🎯 Goal First: What are we trying to do?

We want a network to compare two things and learn whether they are similar or dissimilar.

- Show it pairs of inputs
  - Label them as:
    - 0 = similar
    - 1 = dissimilar
- Let the network learn embeddings (i.e., numeric summaries) for each input
- Use a loss function to:
  - Pull embeddings together if similar
  - Push them apart if dissimilar

🧠 Step-by-Step Intuition: Deriving Contrastive Loss

🧩 Step 1: Define a distance between the two embeddings

Let's say:

- The network creates embedding A for input 1
- Embedding B for input 2
- We calculate the distance D between A and B (using Euclidean, cosine, etc.)
- Think of this like the distance between two dots on a 2D map.
- Now let's ask...

🤝 Step 2: What should happen for similar pairs (label = 0)?

If two things are similar, we want their distance to be:

✅ As small as possible (close together)

So, the loss function should:

$$\text{Loss} = D^2 \quad \text{(for similar pairs)}$$

This way:

- If the distance is large → big penalty
- If distance is small → small penalty

> 📌 Intuition: You're **rewarding closeness** for similar pairs.

✋ Step 3: What should happen for dissimilar pairs (label = 1)?

✅ At least some margin apart (say, 1 or 2 units)

❌ But we don’t want to push them apart forever (no need to separate the moon from the earth!)

So we use:

$$\text{Loss} = \max(0, \text{margin} - D)^2 \quad \text{(for dissimilar pairs)}$$

This means:

- If they are **already far enough apart** → loss = 0 (no problem!)
- If they are **too close** → penalize!

> 📌 Intuition: You're telling the network:
> "These are supposed to be different — make sure they don’t look too close!"

🧮 Step 4: Combine both cases using the label Y

Let's say:

- $Y = 0$ for similar
- $Y = 1$ for dissimilar

Then we write a **single formula** to handle both:

$$\text{Contrastive Loss} = (1 - Y) \cdot D^2 + Y \cdot \max(0, \text{margin} - D)^2$$

✅ Final Intuition Recap

Case	Label (Y)	Desired Action	Loss Term
Similar	0	Pull together	D^2
Dissimilar	1	Push apart (at least by margin)	$(\text{margin} - D)^2$ if D < margin

🧮 Step-by-Step Numerical Example

Case 1: Similar Pair (Y = 0)

Let's say:

- Embedding distance: $D = 0.5$
- Label $Y = 0$ (similar pair)
- Margin $m = 1$

Plug into formula:

$$L = (1 - 0) \cdot (0.5)^2 + 0 \cdot \max(0, 1 - 0.5)^2$$

$$L = 1 \cdot 0.25 + 0 \cdot 0.25 = 0.25$$

✅ So, the loss is **0.25**.

👉 This means: There's still a little distance between the similar pair, so the model gets a small penalty.

Case 2: Dissimilar Pair, Too Close (Y = 1)

Let’s say:

- $D = 0.3$ (too close!)
- $Y = 1$
- $m = 1$

Now:

$$L = (1 - 1) \cdot (0.3)^2 + 1 \cdot \max(0, 1 - 0.3)^2$$

$$L = 0 + 1 \cdot (0.7)^2 = 0.49$$

✅ Loss = **0.49** → Higher penalty

👉 Because the dissimilar pair is not far enough apart (they’re only 0.3 units away instead of at least 1).

Case 3: Dissimilar Pair, Far Enough (Y = 1)

Let's say:

- $D = 1.5$
- $Y = 1$
- $m = 1$

Now:

$$L = (1-1) \cdot (1.5)^2 + 1 \cdot \max(0, 1 - 1.5)^2$$

$$L = 0 + 1 \cdot (0)^2 = 0$$

✅ Loss = **0**

👉 No penalty! The dissimilar pair is far enough apart — they're already more than the required margin.

🎯 Summary

Case	Y (Label)	Distance (D)	Margin (m)	Loss	Explanation
Similar Pair (close)	0	0.5	1	0.25	Still a bit apart, small penalty
Dissimilar Pair (too close)	1	0.3	1	0.49	Not far enough, medium penalty
Dissimilar Pair (far enough)	1	1.5	1	0	Already separated, no penalty