



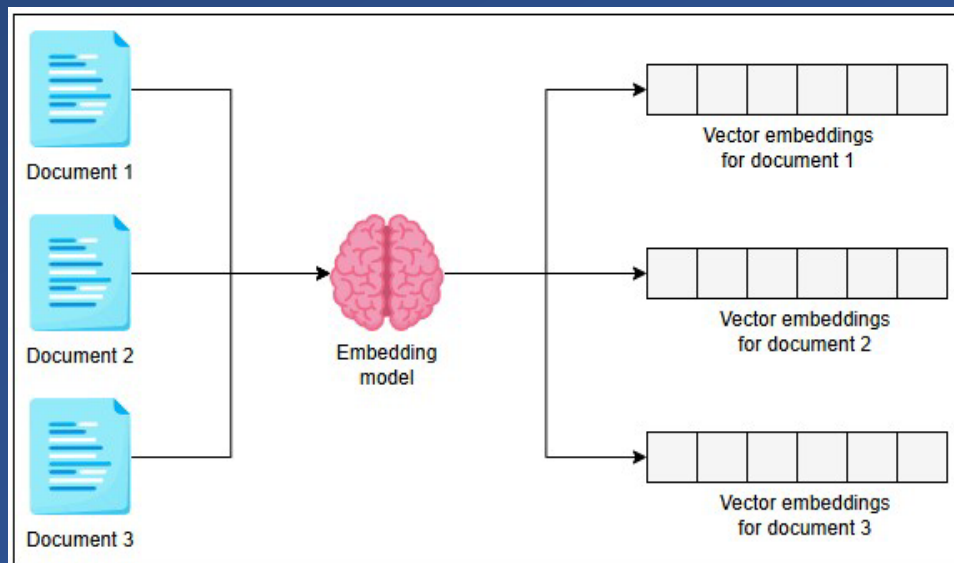
Embedding



גלעד מרקמן
[פרויקט ב-Github](#)

מה זה Embedding ?

- Embedding - דרך לייצג טקסטים (מילים, משפטים, מסמכים) כווקטורים מספריים (מערך מספרים).
- כל טקסט הופך לרשימה ארוכה של מספרים (למשל 1,536 מספרים).
- ההמרה נעשית על ידי מודל של בינה מלאכותית.
- המספרים מייצגים את משמעות הטקסט.

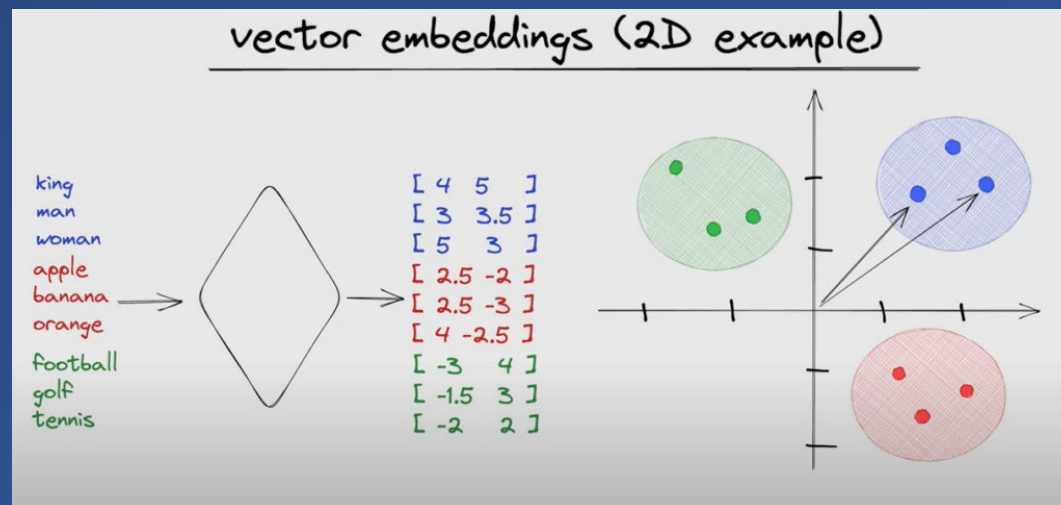


כיצד מבצעים Embedding

- מודלים של בינה מלאכותית (רשתות נוירונים) שאומנו על כמויות עצומות של טקסט, יודעים "לתרגם" טקסט לווקטורים.
- לכל מודל יש גודל וקטור קבוע, למשל:
 - מודל: OpenAI text-embedding-3-small -> 1,536 מימדים (ערכים).
 - מודל: OpenAI text-embedding-3-large -> 3,072 מימדים (ערכים).
 - מודל: HuggingFace Sentence-BERT (SBERT) -> 768 מימדים (קוד פתוח).
- אותו טקסט שיוזן לשני מודלים שונים יניב וקטורים באורך שונה.
- גם הערכים עצמם שונים ממודל למודל. כל מודל למד ייצוג אחר.
- המודל מוציא רשימת מספרים (Floating Point Values) שמתארת את המשמעות הסמנטית של הטקסט.

המטרה של Embeddings

- מאפשר למחשב להבין משמעות, לא רק מילים.
- שימושים מרכזיים:
 - חיפוש סמנטי (Semantic Search).
 - קיבוץ טקסטים לפי נושאים (Clustering).
 - סיווג מסמכים (Classification).
- חיפוש סמנטי מאפשר לנו לשלב ידע חיצוני עם LLM, אלגוריתם המכונה מכונה RAG.

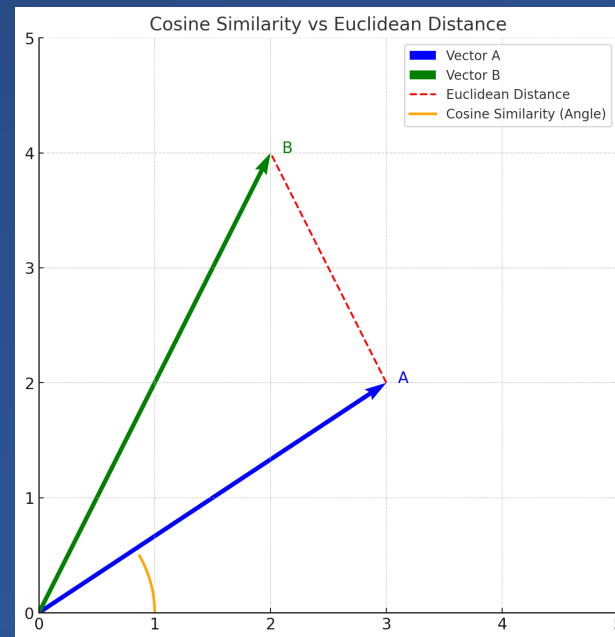
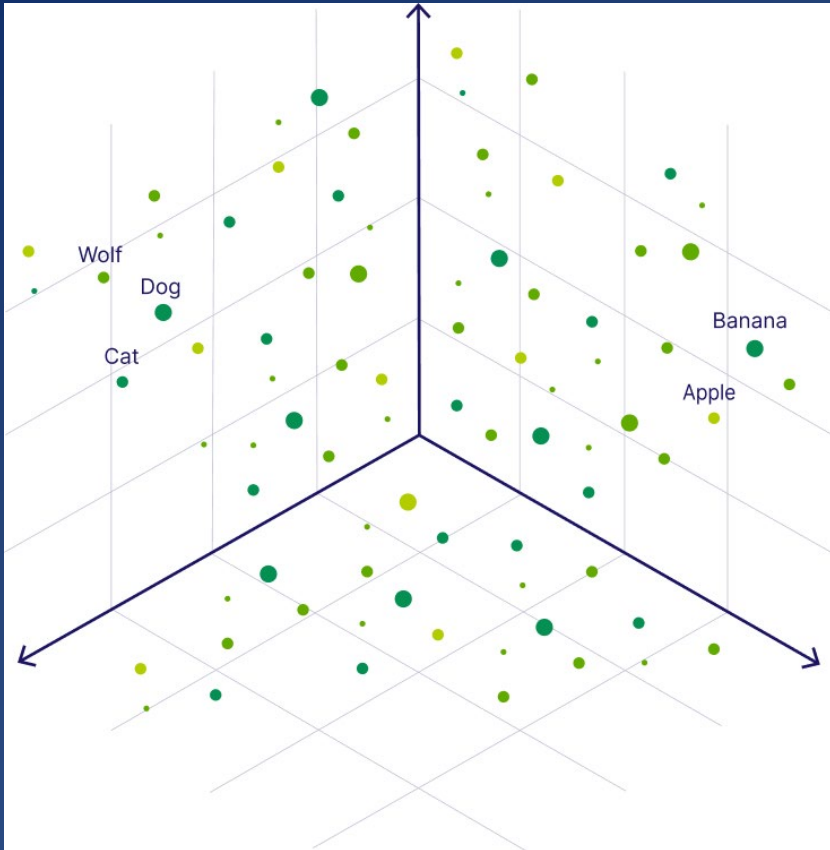


כיצד מודדים דמיון מילולי באמצעות וקטורים

- ניתן למדוד דמיון מילולי של וקטורים באמצעות שני חישובים:

- Cosine Similarity – מדידת כיוון (זווית) של הווקטור.

- Distance Similarity – מדידת מרחק גאומטרי (אוקלידי).



Cosine Similarity

הנוסחה לחישוב Cosine Similarity בין שני וקטורים A, B היא:

$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$

$A \cdot B = \text{Dot Product}$

$\|A\| = \text{Length (Magnitude)}$

דוגמה מספרית

$$A = (3, 4); B = (4, 2)$$

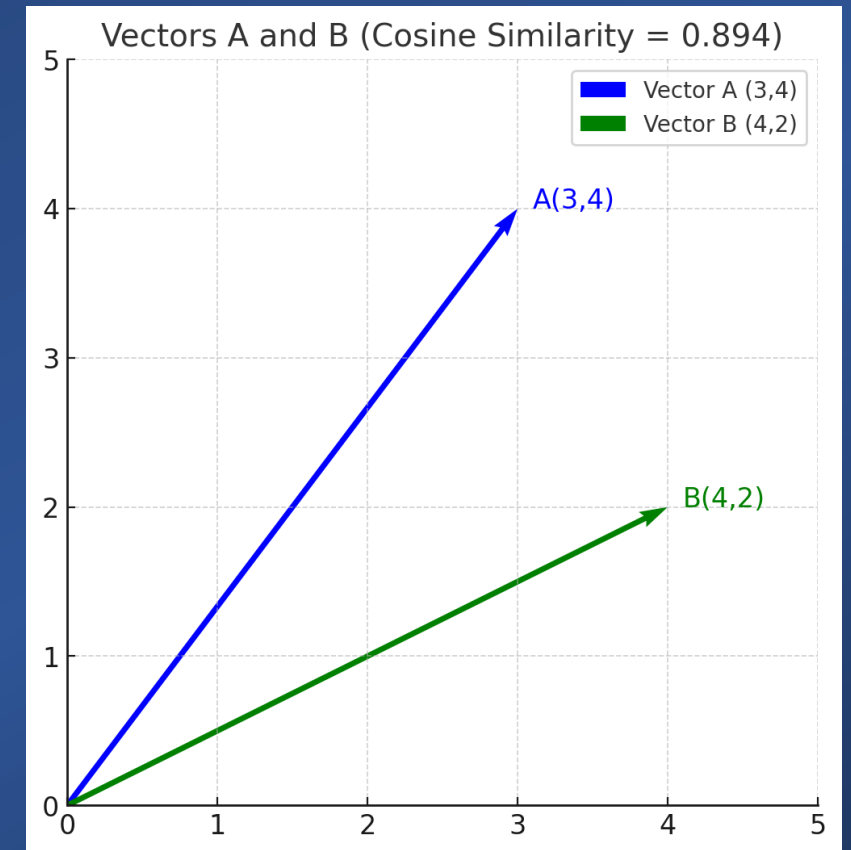
$$A \cdot B = (3 * 4) + (4 * 2) = 12 + 8 = 20$$

$$\|A\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$

$$\|B\| = \sqrt{4^2 + 2^2} = \sqrt{20} = 4.472$$

$$\cos(A, B) = \frac{20}{5 * 4.472} = 0.894$$

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$



דוגמה נוספת

$$A = (1,2,3); B = (4,5,6)$$

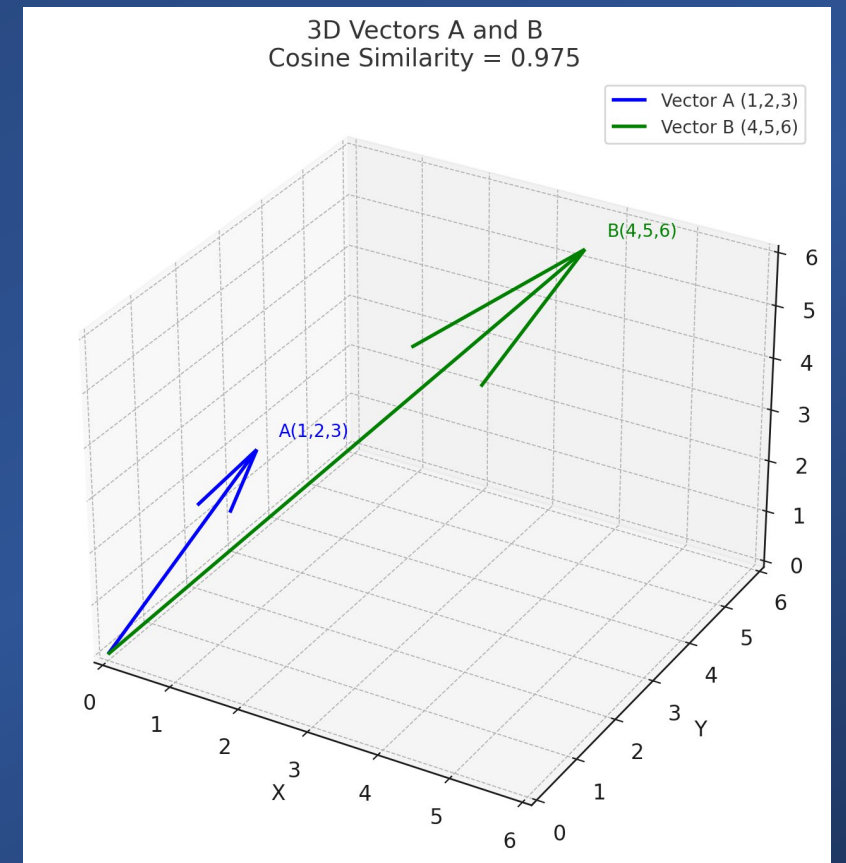
$$A \cdot B = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32$$

$$\|A\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14} = 3.742$$

$$\|B\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77} = 8.775$$

$$\cos(A, B) = \frac{32}{3.742 * 8.775} = 0.975$$

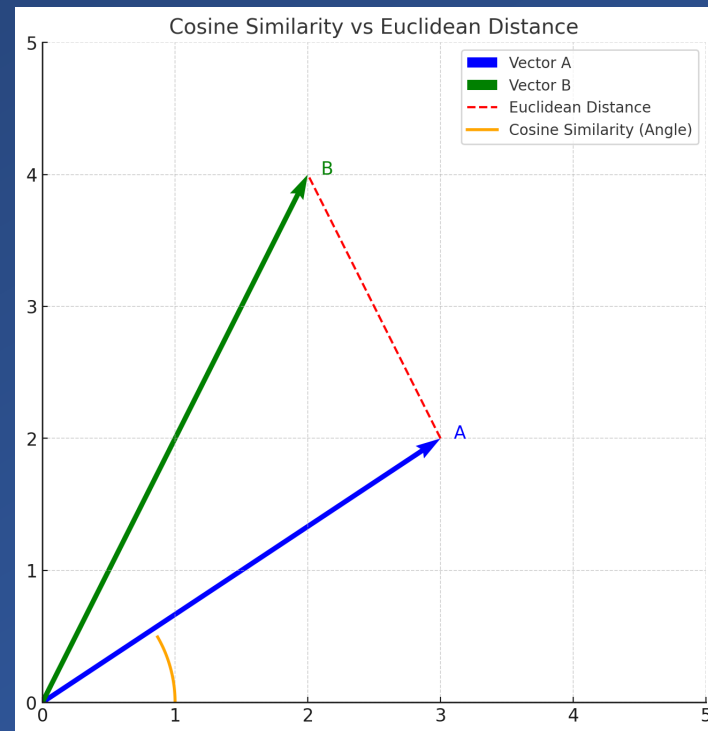
$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$



Distance Similarity

הנוסחה לחישוב Distance Similarity בין שני וקטורים A, B היא:

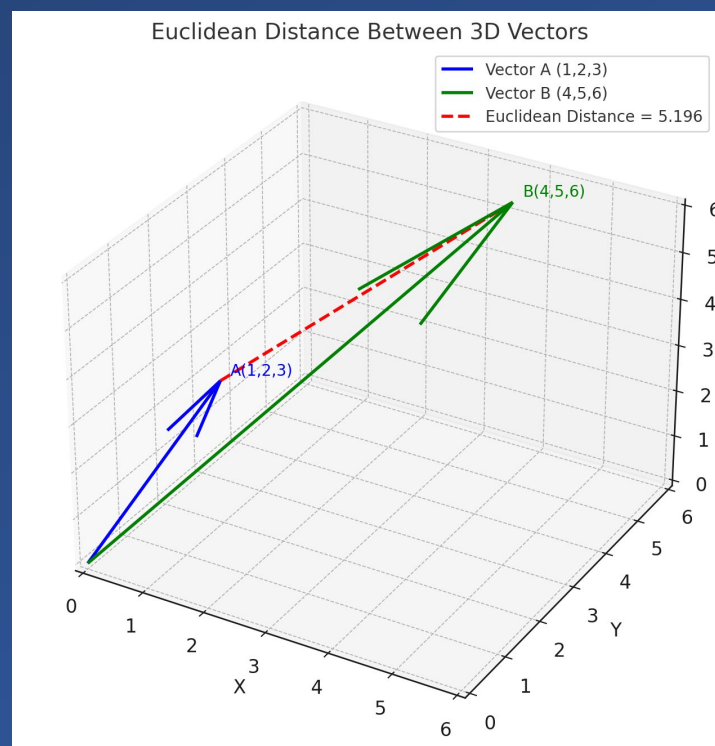
$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$



דוגמה מספרית

$$A = (1,2,3); B = (4,5,6)$$

$$d(A,B) = \sqrt{(1-4)^2 + (2-5)^2 + (3-6)^2} = \sqrt{27} \\ = 5.196$$



Class EmbeddingDemo

נדגים את הנושא באמצעות המחלקה EmbeddingDemo.
נשתמש במודל OpenAI text-embedding-3-small.

```
// Create AI kernel
_kernel = Kernel.CreateBuilder()
    .AddOpenAIEmbeddingGenerator(_embeddingModel, openAiKey)
    .Build();

_embeddingGenerator = _kernel.GetRequiredService<IEmbeddingGenerator<string, Embedding<float>>>();
```

```
// Convert text to numbers (embeddings) using AI
var embedding = await _embeddingGenerator.GenerateAsync(paragraph);
return embedding;
```

נשתמש ב CosineSimilarity אותו נחשב בעצמנו.
נדגים על מספר טקסטים.