PREPARAÇÃO PARA O CURSO DE LARGE LANGUAGE MODELS (LLMs)

Parte 1 de 3

Vetores

Listas de números que descrevem o mundo — e as palavras

Ao final desta parte, você vai conseguir:

- explicar o que é um vetor usando exemplos do cotidiano;
- somar vetores, multiplicá-los por um número e calcular seu tamanho (norma);
- calcular o produto escalar e usá-lo para medir semelhança;
- entender, de forma intuitiva, como um LLM transforma palavras em vetores (embeddings).

Como usar este material

Leia cada seção com calma, refaça as contas dos exemplos com lápis e papel e só depois vá para os exercícios. As caixas 🔗 **Ponte para LLMs** mostram onde cada ideia vai reaparecer no curso — não se preocupe em entender tudo delas agora; elas são um "aperitivo". O gabarito fica no final do documento.

Evandro Barros - evandrobarros.com

1. O que é um vetor?

Um **vetor (vector)** é simplesmente uma **lista ordenada de números**. Só isso. Cada número descreve uma característica de alguma coisa, e a posição de cada número na lista tem um significado fixo.

Pense na ficha de um café na lanchonete. Se combinarmos que a lista vai sempre seguir a ordem **[preço em R$, doçura de 0 a 10, cafeína de 0 a 10]**, então:

Bebida	Vetor	Leitura
Expresso	`[5, 1, 9]`	R$ 5, quase sem açúcar, muita cafeína
Cappuccino	`[9, 6, 6]`	R$ 9, docinho, cafeína média
Chocolate quente	`[8, 9, 0]`	R$ 8, bem doce, sem cafeína

> ⚠️ **Cuidado — a ordem importa**
>
> `[5, 1, 9]` e `[9, 1, 5]` são vetores **diferentes**. No primeiro, o café custa R$ 5 e tem cafeína 9; no segundo, custa R$ 9 e tem cafeína 5. Quem cria o vetor precisa combinar o significado de cada posição — e manter esse combinado sempre.

1.1 Dimensão

A quantidade de números no vetor é a sua **dimensão (dimension)**. O vetor do café tem dimensão 3. Outros exemplos do dia a dia:

- **Localização no mapa:** `[latitude, longitude]` → dimensão 2. Ex.: Barrie ≈ `[44.39, -79.69]`.
- **Dados de saúde de um app de celular:** `[passos, horas de sono, batimentos em repouso]` → dimensão 3.
- **Cor de um pixel na tela:** `[vermelho, verde, azul]`, cada um de 0 a 255 → dimensão 3. Amarelo puro é `[255, 255, 0]`.
- **Boletim de um aluno:** `[Português, Matemática, História, Ciências, Inglês]` → dimensão 5.

> 💡 **Dica — ponto ou vírgula?**
>
> Dentro dos vetores usamos o **ponto** como separador decimal, como em programação: `[1.5, 0.5]` quer dizer "um e meio" e "meio". Assim a vírgula fica reservada para separar os elementos da lista. No texto corrido, seguimos o padrão brasileiro (6,48).

Não existe limite: um vetor pode ter 2, 50 ou 4.096 números. Nós não conseguimos *desenhar* mais de 3 dimensões, mas a matemática funciona exatamente igual.

2. Vetor como seta: direção e tamanho

Vetores com 2 ou 3 números podem ser desenhados como uma **seta**. Imagine que, para ir de casa até a padaria, você anda **3 quarteirões para o leste e 4 para o norte**. Esse deslocamento é o vetor `[3, 4]`.

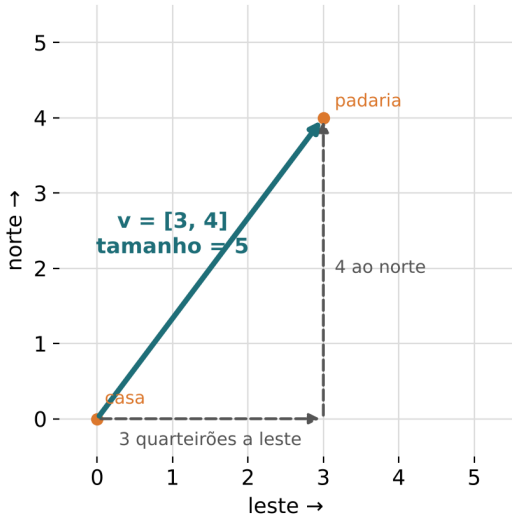


Figura 1 — O deslocamento casa → padaria representado como o vetor [3, 4].

O **tamanho** da seta (a distância em linha reta) se chama **norma (norm)** ou **magnitude (magnitude)**. Calculamos com o bom e velho Pitágoras: elevamos cada número ao quadrado, somamos e tiramos a raiz quadrada.

$$\|[3,\ 4]\| = \sqrt{(3^2 + 4^2)} = \sqrt{(9 + 16)} = \sqrt{25} = 5$$

Ou seja: andando em linha reta (se fosse possível atravessar os prédios), a padaria fica a 5 quarteirões. A mesma conta vale para qualquer dimensão: para `[1, 2, 2]`, a norma é $\sqrt{(1 + 4 + 4)} = \sqrt{9} = 3$.

3. Operações básicas com vetores

3.1 Soma de vetores (vector addition)

Somamos vetores **posição por posição**. Exemplo: as compras da semana, na ordem **[maçãs, pães, litros de leite]**:

```
Segunda [2, 1, 3]  +  Terça [1, 2, 0]  =  Total [3, 3, 3]
```

Só dá para somar vetores de **mesma dimensão** e **mesmo significado** em cada posição. Somar `[maçãs, pães]` com `[gols, faltas]` não faz sentido nenhum.

3.2 Multiplicação por um número (escalar)

Um número sozinho, sem ser lista, se chama **escalar (scalar)**. Multiplicar um vetor por um escalar multiplica **cada posição** por esse número. É exatamente o que fazemos ao **dobrar uma receita**:

```
Bolo para 4 pessoas: [200 g farinha, 2 ovos, 100 ml leite]

2 × [200, 2, 100] = [400, 4, 200]   → bolo para 8 pessoas
```

Repare que as proporções entre os ingredientes não mudam — só a "escala". Na seta, multiplicar por 2 deixa a seta duas vezes mais comprida, apontando para o mesmo lado.

3.3 Produto escalar (dot product) — a operação mais importante deste material

O **produto escalar (dot product)** pega dois vetores do mesmo tamanho, **multiplica posição por posição e soma tudo**. O resultado é um único número.

Você já faz isso no caixa do supermercado! Quantidades `[3, 2, 1]` (3 iogurtes, 2 queijos, 1 café) e preços `[4, 10, 6]`:

```
[3, 2, 1] · [4, 10, 6] = 3×4 + 2×10 + 1×6 = 12 + 20 + 6 = R$ 38
```

O total da compra é o produto escalar entre o vetor de quantidades e o vetor de preços. Guarde esta ideia: **"multiplica par a par e soma"**. Ela vai aparecer o tempo todo — em matrizes, em redes neurais e no coração dos LLMs.

4. Medindo semelhança entre vetores

Aqui o produto escalar ganha um superpoder: ele indica **o quanto dois vetores "apontam para o mesmo lado"**. Quando as posições grandes de um coincidem com as posições grandes do outro, o resultado é alto.

Exemplo: quatro amigos deram nota de 0 a 5 para filmes de **[ação, romance, comédia]**. Vamos comparar cada um deles com a Ana:

Pessoa	Gosto (vetor)	Produto escalar com Ana
Ana	`[5, 1, 4]`	5×5 + 1×1 + 4×4 = **42** (com ela mesma)
Bruno	`[4, 0, 5]`	5×4 + 1×0 + 4×5 = **40**
Carla	`[0, 5, 1]`	5×0 + 1×5 + 4×1 = **9**
Edu	`[5, 5, 5]`	5×5 + 1×5 + 4×5 = **50**

Ana e Bruno gostam das mesmas coisas (ação e comédia) → produto escalar alto. Ana e Carla têm gostos bem diferentes → produto escalar baixo. Um serviço de streaming poderia recomendar para a Ana os filmes que o Bruno adorou.

Mas repare no **Edu**: ele dá nota 5 para todo filme, e o produto escalar dele com a Ana deu **50**, mais que os 40 do Bruno! Pela conta, o Edu seria quem mais combina com a Ana. Isso faz sentido? A próxima seção resolve esse mistério.

4.1 Similaridade de cosseno (cosine similarity)

O produto escalar tem um problema: ele fica grande não só quando os gostos combinam, mas também quando as notas são **altas**. É o caso do Edu: ele não tem preferência nenhuma (dá 5 para ação, romance e comédia), só é generoso nas notas. Mesmo assim, ganhou do Bruno. O produto escalar mistura duas coisas: **o gosto** (para onde a seta aponta) e **a intensidade** (o tamanho da seta).

Queremos comparar só o gosto. Pense nas receitas da seção 3.2: o bolo para 4 pessoas `[200, 2, 100]` e o bolo para 8 `[400, 4, 200]` têm quantidades diferentes, mas são **a mesma receita**, porque as proporções são iguais. Queremos comparar as proporções (a direção da seta), e não as quantidades (o tamanho da seta). Fazemos isso em dois passos.

Passo 1 — medir o tamanho (norma) de cada seta. É o Pitágoras da seção 2: elevamos cada nota ao quadrado, somamos e tiramos a raiz quadrada. A notação ‖Ana‖ quer dizer apenas "tamanho do vetor da Ana".

Pessoa	Notas	Quadrados	Soma	Tamanho (raiz)
Ana	`[5, 1, 4]`	25 + 1 + 16	42	√42 ≈ **6,48**
Bruno	`[4, 0, 5]`	16 + 0 + 25	41	√41 ≈ **6,40**
Carla	`[0, 5, 1]`	0 + 25 + 1	26	√26 ≈ **5,10**
Edu	`[5, 5, 5]`	25 + 25 + 25	75	√75 ≈ **8,66**

Repare que o Edu tem a seta mais comprida: é o efeito das notas altas.

Passo 2 — dividir o produto escalar pelos dois tamanhos. Assim "descontamos" o tamanho das setas e sobra só a comparação de direção. O resultado se chama **similaridade de cosseno (cosine similarity)**:

$$\text{cosseno}(A, B) = (A \cdot B) \div (\|A\| \times \|B\|)$$

Comparação	Produto escalar (multiplica nota por nota e soma)	Dividido por (tamanho × tamanho)	Cosseno
Ana e Bruno	5×4 + 1×0 + 4×5 = **40**	6,48 × 6,40 ≈ 41,5	40 ÷ 41,5 ≈ **0,96**
Ana e Edu	5×5 + 1×5 + 4×5 = **50**	6,48 × 8,66 ≈ 56,1	50 ÷ 56,1 ≈ **0,89**
Ana e Carla	5×0 + 1×5 + 4×1 = **9**	6,48 × 5,10 ≈ 33,0	9 ÷ 33,0 ≈ **0,27**

Agora o resultado faz sentido. O Bruno volta ao primeiro lugar (0,96): ele gosta das mesmas coisas que a Ana. O Edu cai para o segundo (0,89): parece com a Ana só porque gosta de tudo, inclusive de romance, que ela não curte. A Carla fica bem longe (0,27). E as duas receitas de bolo dariam cosseno exatamente **1**: mesma direção, só tamanhos diferentes.

Outra vantagem: o cosseno é sempre um número entre −1 e 1, quer as notas sejam de 0 a 5 ou de 0 a 1.000. Isso permite comparar qualquer par de vetores na mesma régua.

> 💡 **Dica — como interpretar o cosseno**
>
> - Perto de **1** → muito parecidos (setas apontando para o mesmo lado).
> - Perto de **0** → sem relação (setas formando ângulo de 90°).
> - Perto de **−1** → opostos (setas apontando para lados contrários).

5. Ponte para LLMs: palavras viram vetores

Um computador não entende a palavra "gato". Ele só entende números. A grande ideia por trás dos modelos de linguagem é transformar cada palavra (ou pedaço de palavra, o **token (token)**) em um vetor. Esse vetor se chama **embedding (embedding)**, ou "vetor de representação".

Vamos inventar embeddings bem simples, com 3 dimensões que nós mesmos escolhemos: **[realeza, masculino, comida]**:

Palavra	Embedding (de brinquedo)
rei	`[0.9, 0.9, 0.0]`
rainha	`[0.9, 0.1, 0.0]`
homem	`[0.1, 0.9, 0.0]`
mulher	`[0.1, 0.1, 0.0]`
pizza	`[0.0, 0.5, 0.9]`

Agora repare em duas coisas incríveis:

- **Significado vira distância.** Faça as contas: cos(rei, rainha) = 0,90 ÷ (1,27 × 0,91) **≈ 0,78** (ambos são realeza); cos(rei, pizza) = 0,45 ÷ (1,27 × 1,03) **≈ 0,34**. Palavras com sentido parecido ficam **perto** no espaço de vetores.
- **Relações viram setas.** Faça a conta, posição por posição: rei − homem + mulher = [0.9−0.1+0.1, 0.9−0.9+0.1, 0−0+0] = `[0.9, 0.1, 0.0]` = **rainha**! A "seta" que leva de homem para mulher é a mesma que leva de rei para rainha.

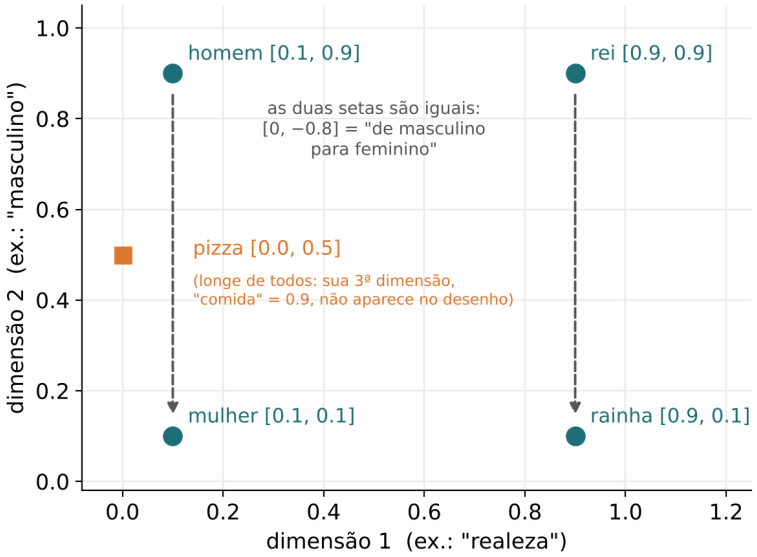


Figura 2 — Palavras como pontos: sentidos parecidos ficam próximos; relações parecidas viram setas parecidas.

🔗 Ponte para LLMs — o que muda nos modelos reais

Em um LLM de verdade, cada token tem um embedding com **centenas ou milhares de dimensões** (768, 4.096, 12.288...). E ninguém escolhe à mão o que cada posição significa: o modelo **aprende** esses números sozinho, lendo bilhões de textos. Mesmo assim, as duas propriedades acima continuam valendo.

Quando o modelo decide "quais palavras da frase têm a ver com quais" (o mecanismo de **atenção (attention)**, que você verá no curso), ele está calculando... **produtos escalares entre vetores**. Exatamente a conta do supermercado.

📌 Resumo

- Vetor = lista ordenada de números; cada posição tem um significado.
- Dimensão = quantos números o vetor tem.
- Norma = tamanho da seta (Pitágoras).
- Soma e multiplicação por escalar = posição por posição.
- Produto escalar = multiplica par a par e soma → mede alinhamento.
- Similaridade de cosseno = produto escalar "normalizado", de −1 a 1.
- Embedding = vetor que representa uma palavra/token; sentidos parecidos ficam próximos.

Exercícios

Resolva com lápis e papel. Calculadora é permitida para raízes quadradas.

Exercício 1 · montar vetores

Combinamos que a ficha de um celular é o vetor **[preço em R$ mil, bateria em horas, câmera em megapixels]**.

a) Escreva o vetor de um celular que custa R$ 2 mil, tem 20 horas de bateria e câmera de 50 MP.
b) Qual é a dimensão desse vetor?
c) O vetor `[50, 2, 20]` descreve o mesmo celular? Por quê?

Exercício 2 · soma

No campeonato da escola, a ficha de um jogador por rodada é **[gols, assistências]**. Rodada 1: `[2, 1]`; rodada 2: `[0, 3]`; rodada 3: `[1, 1]`. Qual é o vetor com o total do campeonato?

Exercício 3 · escalar

Uma receita de bolo usa **[3 xícaras de farinha, 2 ovos, 1 xícara de açúcar]**.

a) Escreva o vetor para fazer **meia receita**.
b) Escreva o vetor para fazer **três receitas**.

Exercício 4 · norma

Um entregador de aplicativo saiu do restaurante e andou 6 quarteirões para o leste e 8 para o norte.

a) Escreva o vetor do deslocamento.
b) Qual a distância em linha reta entre o restaurante e o ponto de entrega?

Exercício 5 · produto escalar

Na lanchonete, você pediu **2 sanduíches, 3 sucos e 1 sobremesa**. Os preços são R$ 12 (sanduíche), R$ 6 (suco) e R$ 8 (sobremesa). Escreva os dois vetores e calcule o total da conta usando o produto escalar.

Exercício 6 · semelhança

Notas de 0 a 5 para estilos musicais **[rock, samba, eletrônica]**: Você `[4, 5, 1]`, Pedro `[5, 4, 0]`, Júlia `[1, 0, 5]`.

a) Calcule o produto escalar entre você e Pedro, e entre você e Júlia.
b) Com quem você deveria trocar playlists? Justifique.
c) (Desafio) Calcule a similaridade de cosseno entre você e Pedro.

Exercício 7 · analogia com embeddings

Use embeddings de brinquedo com as dimensões **[é_capital, França, Itália]**: Paris `[1, 1, 0]`, França `[0, 1, 0]`, Itália `[0, 0, 1]`, Roma `[1, 0, 1]`.

a) Calcule **Paris − França + Itália**.

b) O resultado é igual ao vetor de qual palavra? Que "pergunta" essa conta respondeu?

Exercício 8 · conceitual

Responda com suas palavras:

a) Por que um computador precisa transformar palavras em vetores?

b) Se os embeddings de "cachorro" e "gato" têm cosseno 0,85, e os de "cachorro" e "geladeira" têm cosseno 0,05, o que isso diz sobre essas palavras?

c) Verdadeiro ou falso: "um vetor com 4.096 dimensões precisa de uma matemática diferente de um vetor com 3 dimensões".

Gabarito

Resposta 1

a) `[2, 20, 50]`. b) Dimensão 3. c) Não. Pelo nosso combinado, `[50, 2, 20]` seria um celular de R$ 50 mil, 2 horas de bateria e câmera de 20 MP. A ordem define o significado.

Resposta 2

$[2,1] + [0,3] + [1,1] = [3, 5] \rightarrow$ 3 gols e 5 assistências

Resposta 3

a) $0{,}5 \times [3, 2, 1] =$ `[1.5, 1, 0.5]`. b) $3 \times [3, 2, 1] =$ `[9, 6, 3]`.

Resposta 4

a) `[6, 8]`.

b) $\sqrt{(6^2 + 8^2)} = \sqrt{(36 + 64)} = \sqrt{100} = 10$ quarteirões

Resposta 5

$[2, 3, 1] \cdot [12, 6, 8] = 24 + 18 + 8 = R\$\ 50$

Resposta 6

a) Você·Pedro $= 4{\times}5 + 5{\times}4 + 1{\times}0 = 40$ Você·Júlia $= 4{\times}1 + 5{\times}0 + 1{\times}5 = 9$

b) Com o Pedro: o produto escalar é muito maior, porque vocês dois dão notas altas para os mesmos estilos (rock e samba).

c) $\|\text{Você}\| = \sqrt{42} \approx 6{,}48$ $\|\text{Pedro}\| = \sqrt{41} \approx 6{,}40$ $\cos = 40 \div 41{,}50 \approx 0{,}96$

Resposta 7

a) $[1,1,0] - [0,1,0] + [0,0,1] = [1, 0, 1]$

b) É o vetor de **Roma**. A conta respondeu: "Paris está para a França assim como ___ está para a Itália". Tiramos "ser da França" e colocamos "ser da Itália", mantendo "ser capital".

Resposta 8

a) Porque o computador só faz contas com números. Transformando palavras em vetores, conseguimos medir semelhanças, somar, comparar — e é isso que as redes neurais fazem.

b) "Cachorro" e "gato" têm significados próximos (animais de estimação) e aparecem em contextos parecidos; "cachorro" e "geladeira" quase não têm relação.

c) Falso. As operações (soma, escalar, produto escalar, norma, cosseno) são exatamente as mesmas; só há mais posições para somar.