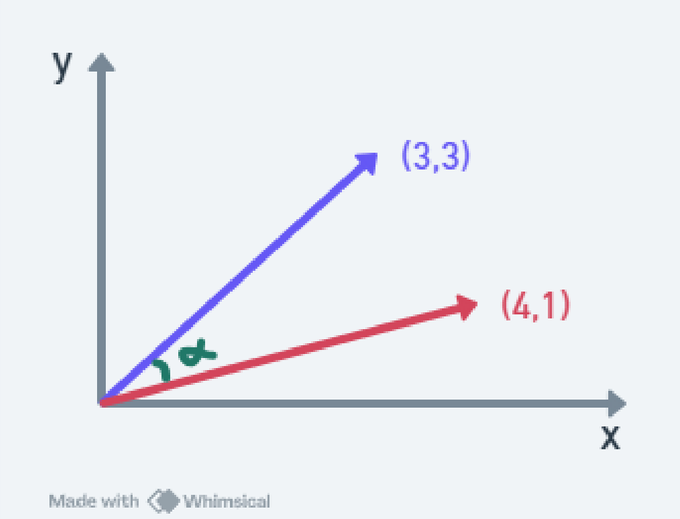
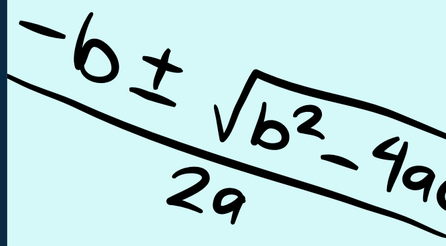
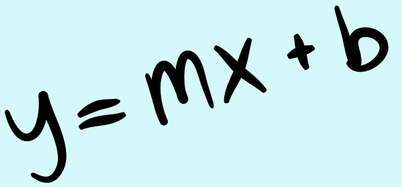
Similaridade do cosseno

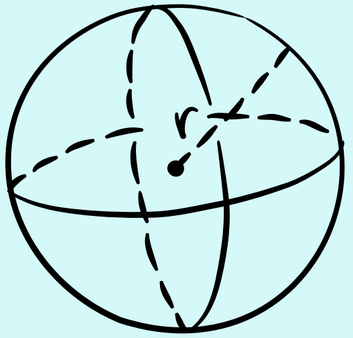
Como funciona

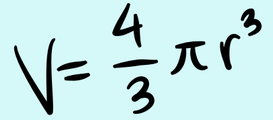
- Mapeamento de vetores dentro de um contexto gráfico
- Calculo do ângulo entre os vetores
- O cosseno do ângulo será a medida de similaridade





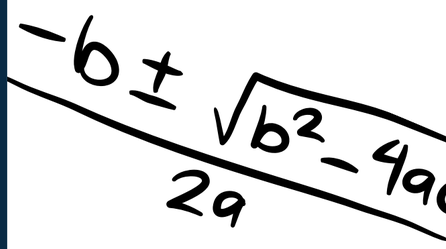


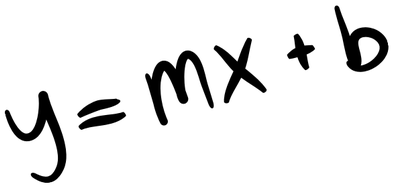


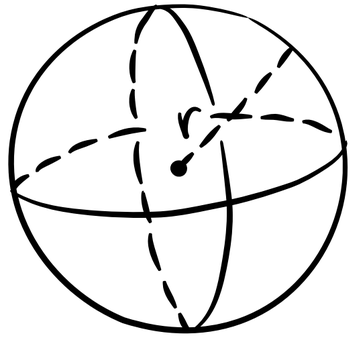


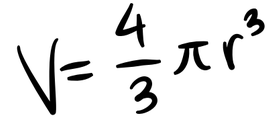
Como funciona

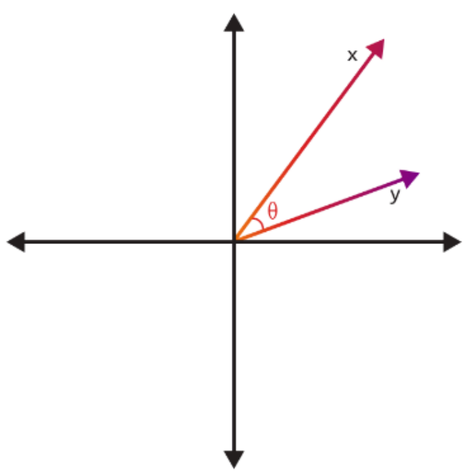
- Ângulo próximo de 0, similaridade próxima de 1
- Ângulo próximo de 90, similaridade próxima de 0
- O resultado é um valor entre -1 e 1
- Quanto maior o valor, mais similar é



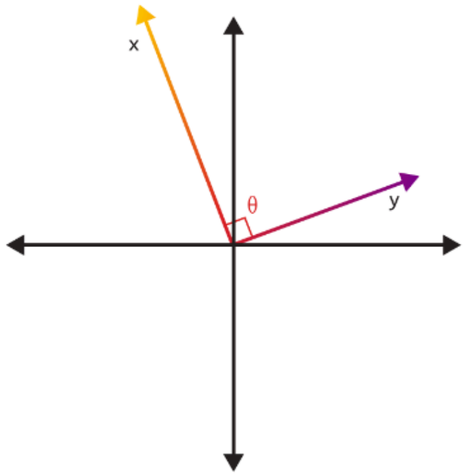




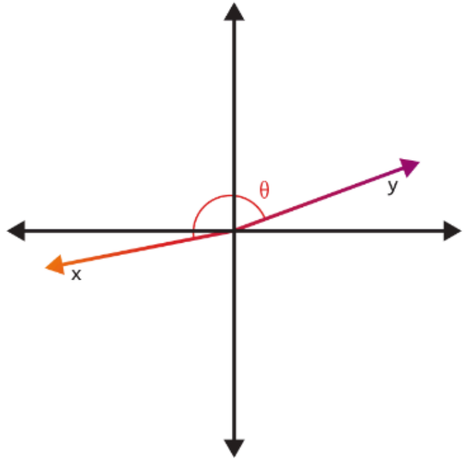




Angle θ close to 0°
Cos(θ) close to 1 → **Similar vectors**



Angle θ close to 90°
Cos(θ) close to 0 → **Orthogonal vectors**

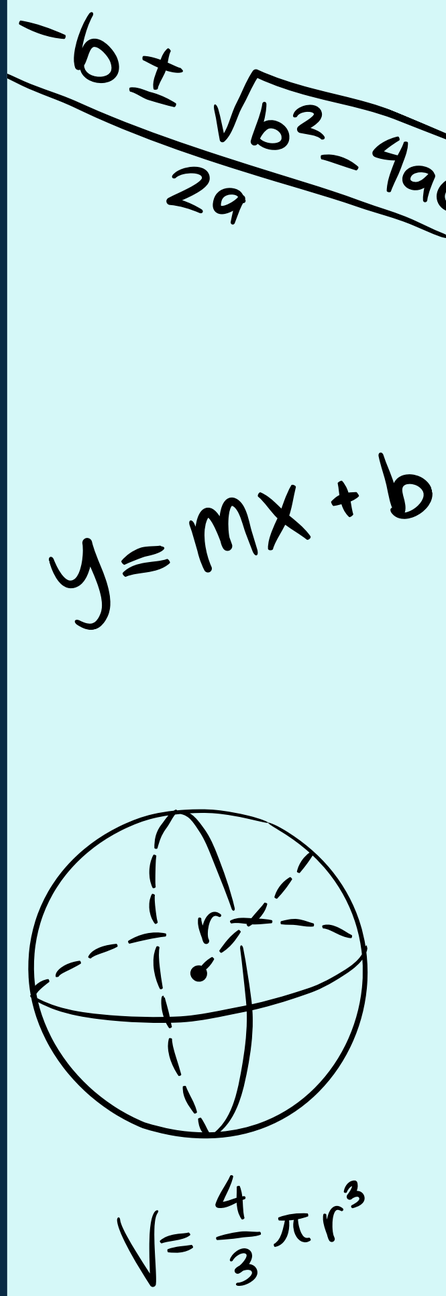


Angle θ close to 180°
Cos(θ) close to -1 → **Opposite vectors**

Fonte: memgraph

USOS

- Similaridade entre documentos
- Clustering
- Classificação de texto



Exemplo

Frase 1

Vamos para escola

Frase 2

Vamos para faculdade

Exemplo

	vamos	para	escola	faculdade
Frase 1	1	1	1	0
Frase 2	1	1	0	1

Exemplo

$$\frac{(A \cdot B)}{(||A|| \,.\, ||B||)} = \frac{(1.1) + (1.1) + (1.0) + (0.1)}{\sqrt{1^2+1^2+1^2+0^2} \,.\, \sqrt{1^2+1^2+0^2+1^2}} = \frac{2}{\sqrt{3} \,.\, \sqrt{3}} = \frac{2}{3} = 0,667$$

EM PYTHON (SCIKIT-LEARN)

```
from sklearn.metrics.pairwise import cosine_similarity

print(cosine_similarity([[1,1,1,0], [1,1,0, 1]]))
```

[1] ✓ 0.5s

```
[[1.         0.66666667]
 [0.66666667 1.        ]]
```

Fontes

Medium

Brains

memgraph