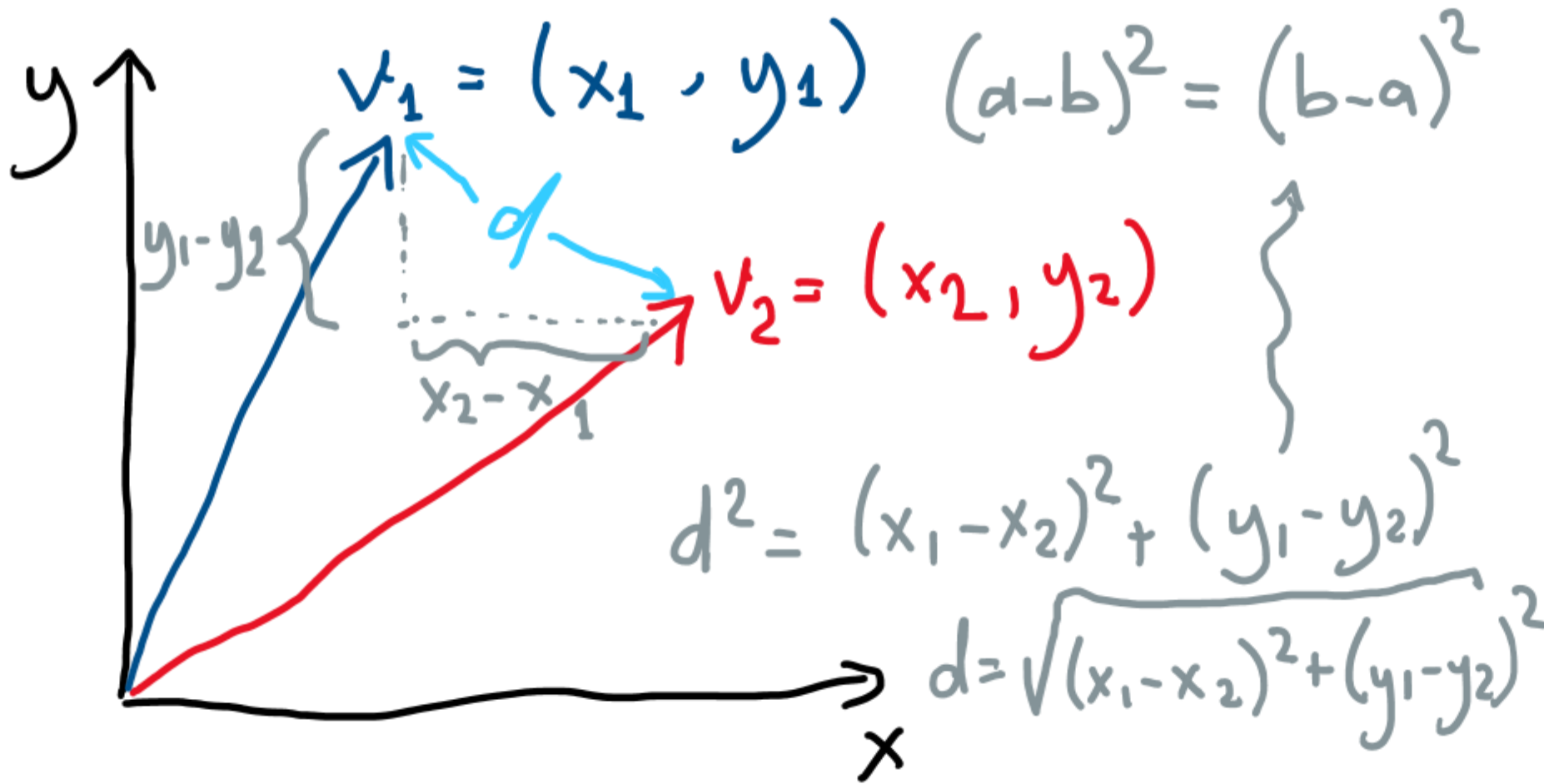


Vektör Benzerlikleri

- d1: ["a", "b", "a", "b", "a", "a", "b"] -> a:4, b:3
 - d2: ["b", "b", "a", "a", "b", "b", "b"] -> a:2, b:5
 - d3: ["b", "a", "a", "b", "b", "a", "a", "a", "b", "b", "a", "a", "b", "a",] -> a:8, b:6
-
- d_i : Dokümanlar
 - Hangi dokümanlar daha benzer?

Euclidean Distance (Öklit Uzaklığı)

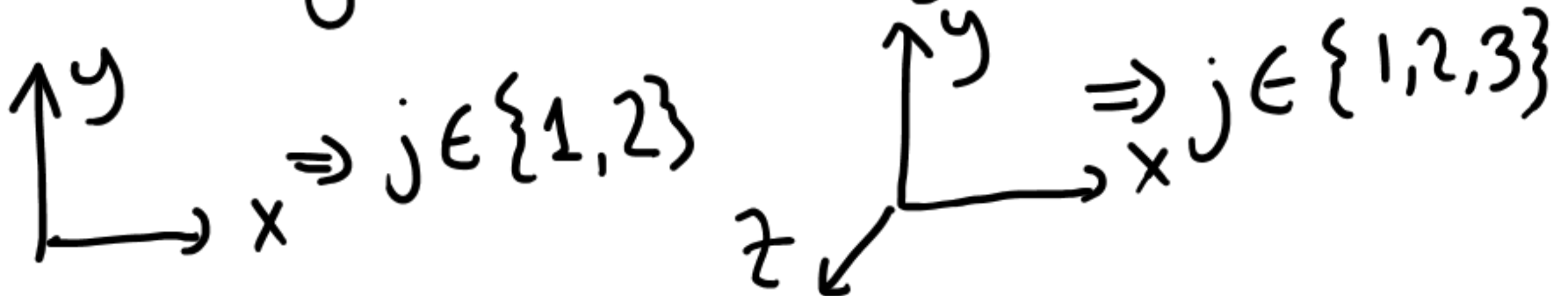


Euclidean Distance

$$\|v_1 - v_2\|_2 = \sqrt{(v_{11} - v_{21})^2 + (v_{12} - v_{22})^2 + \dots + (v_{1D} - v_{2D})^2}$$

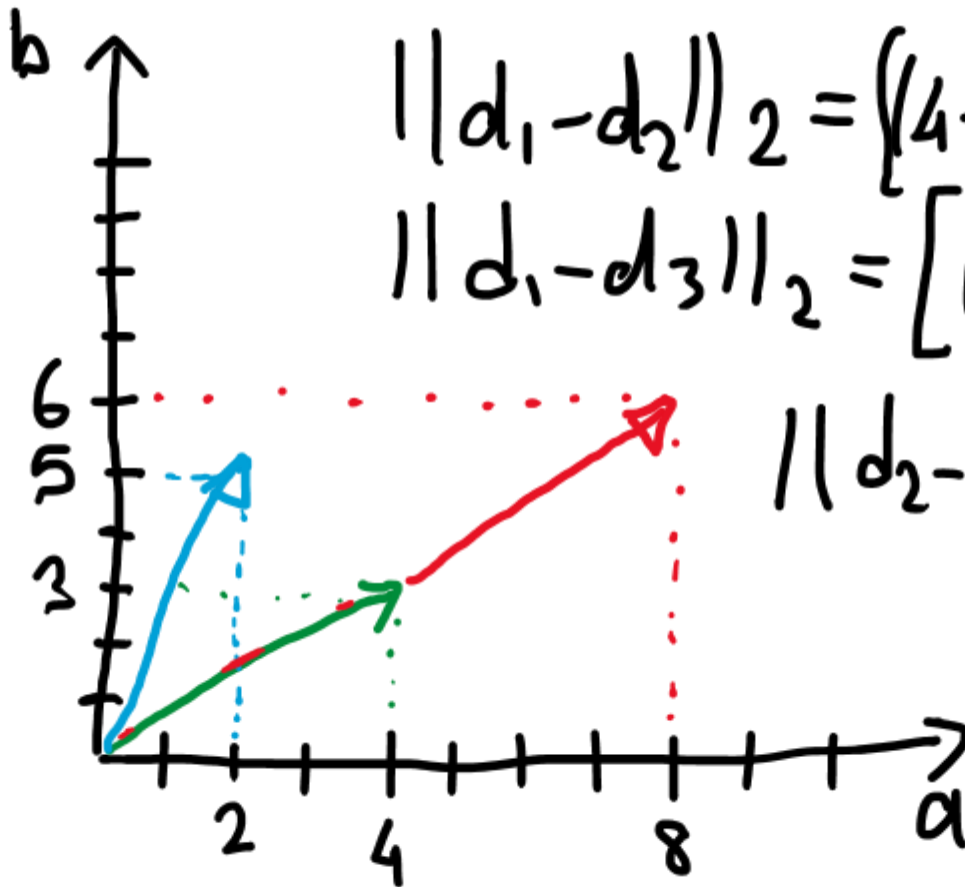
v_{ij} : $i \Rightarrow$ hangi vektör

$j \Rightarrow$ vektörün j 'inci bileşeni



Benzerlik

- d1 -> a:4, b:3; d2 -> a:2, b:5; d3 -> a:8, b:6



$$\|d_1 - d_2\|_2 = [(4-2)^2 + (3-5)^2]^{1/2} = 8^{1/2}$$

$$\|d_1 - d_3\|_2 = [(4-8)^2 + (3-6)^2]^{1/2} = 25^{1/2}$$

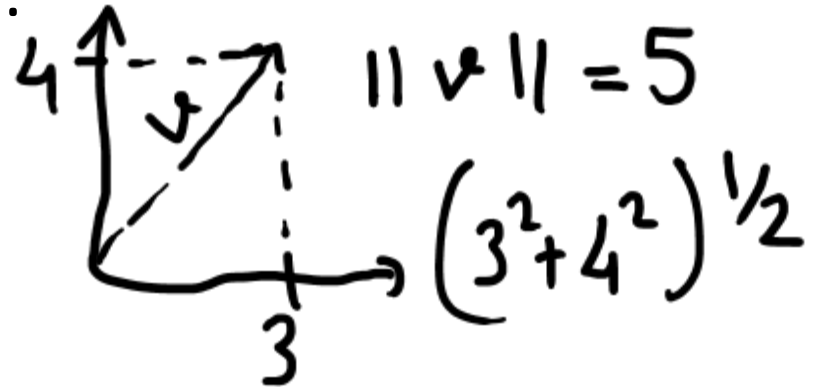
$$\|d_2 - d_3\|_2 = [(2-8)^2 + (5-6)^2]^{1/2} = 37^{1/2}$$

$$\begin{aligned} d_1 - d_2 &: 8 \checkmark \\ d_1 - d_3 &: 25 \\ d_2 - d_3 &: 37 \end{aligned}$$

Çözüm 1: Normalizasyon

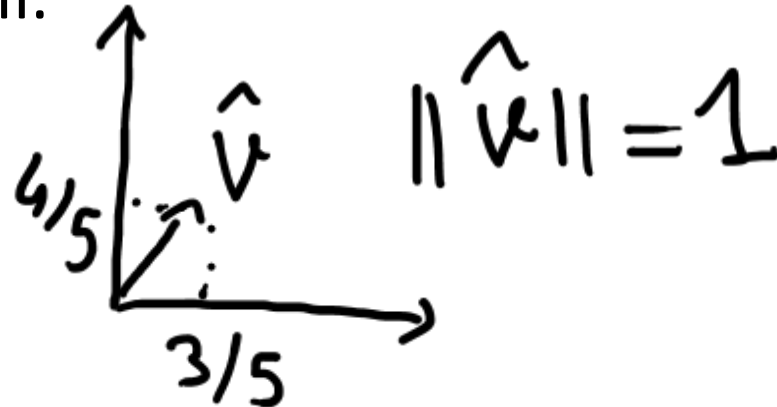
- Vektörlerin boyunu 1'e indirmek.
- D boyutlu bir vektörün boyu:

$$\|v\| = \left(\sum_{j=1}^D v_j^2 \right)^{1/2}$$

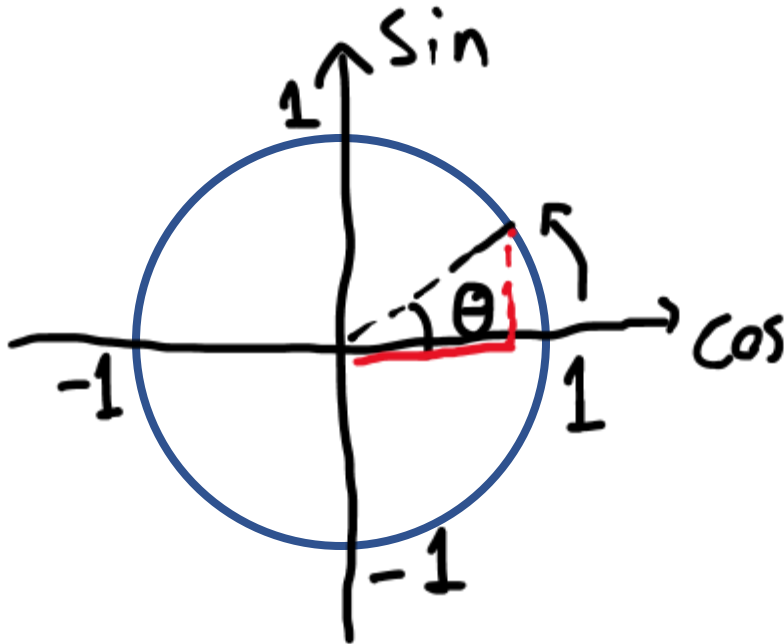


- Normalize etme işlemi:

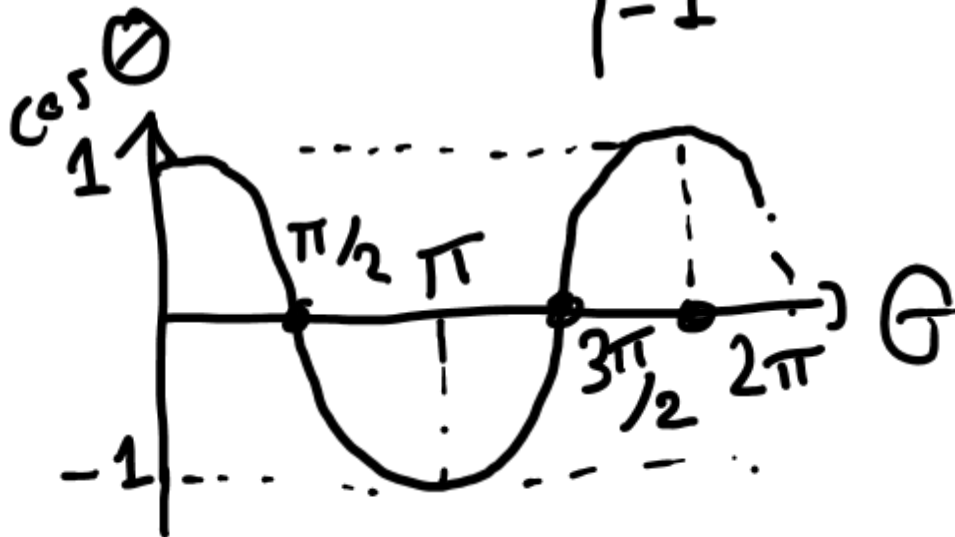
$$\hat{v} = \frac{v}{\|v\|}$$



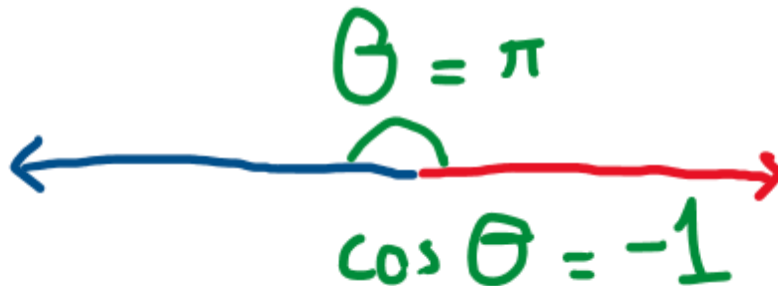
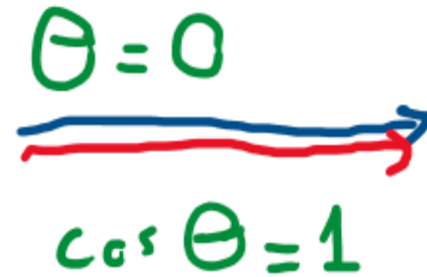
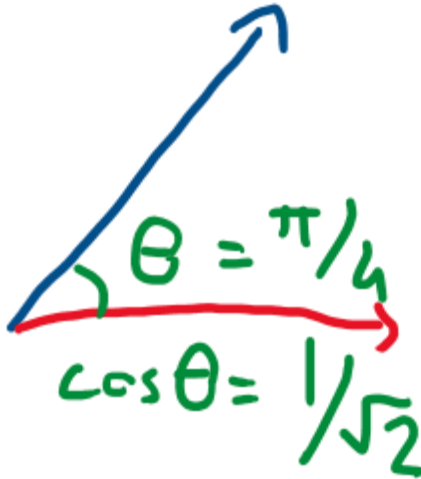
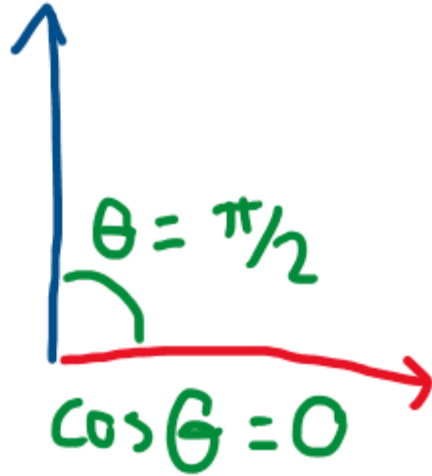
Çözüm 2: Cosine Similarity



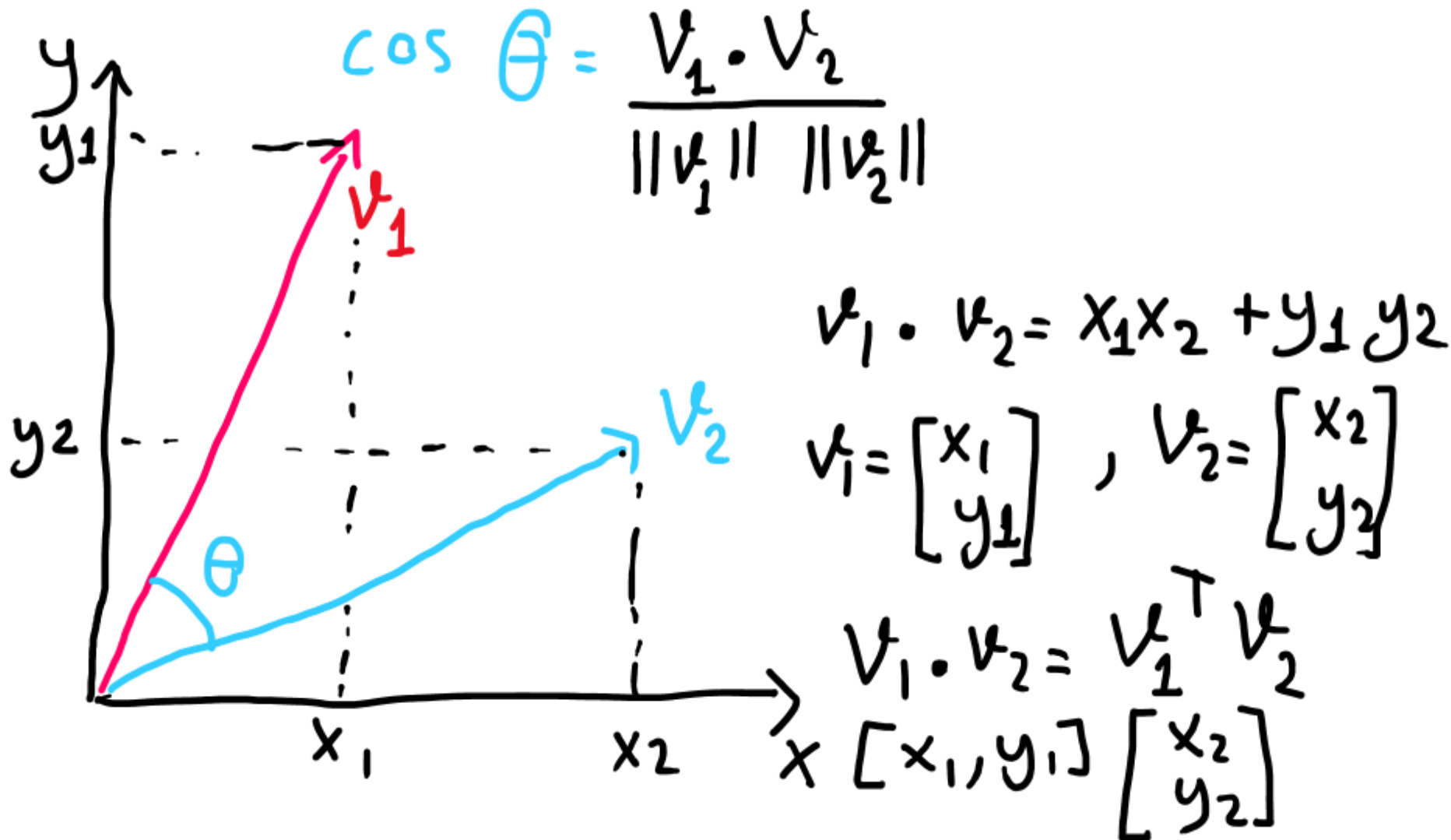
θ	$\cos \theta$
0	1
$\pi/2$	0
π	-1
$3/2\pi$	0
2π	1



Çözüm 2: Cosine Similarity



Çözüm 2: Cosine Similarity



Çözüm 2: Cosine Similarity

