

# THEORY

## Step 1

Movie title	Movie Tag
Inception	Sci-Fi Thriller Dream Heist Leo Hardy..
Interstellar	... Drama Space Future Mc...
The Dark Knight	Action, Crime, Gotham, Joker, Bale;...

## Step 2:

We extract all unique words from these tags to form vocabulary:  
total = 18

## Step 3:

Each movie is now represented as an 18-dimensional Vector, where each dimension corresponds to a word from vocab.

	Action	Bale	Crime	Drama	Dream	Future	Gotham	Hardy	Heist	Joker
Inception	1	0	0	0	1	0	0	1	1	0
Interstellar										
The Dark Knight										

Leo	Ledger	McConaughey	Nolan	Sci-Fi	Space	Thriller	Zimmer
1	0	0	1	1	0	1	1

Like this we will read every sentence and we convert into Vector.



Each row is a vector representation of movie

$$\text{Inception} = [0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1]$$

$$\text{Interstellar} = [0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0]$$

$$\text{The Dark Knight} = [1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1]$$

Step 3: Compute similarity Using cosine similarity

$$\text{similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|}$$

where  $A \cdot B$  is dot product of two vector

$\|A\|$  and  $\|B\|$  are magnitude [Euclidean norms]

It is a function that measure  
the length of a vector in a  
Euclidean space

$\Rightarrow A \cdot B = (a_1 \times b_1) + (a_2 \times b_2) + \dots + (a_n \times b_n)$   
lets compute between Inception vs Interstellar:

$$\Rightarrow (0 \times 0) + (0 \times 0) + \dots$$

$$\Rightarrow 1 + 1 + 1 + 1$$

now for Inception vs The Dark Knight  
 $= 1 + 1 + 1 = 3$



↳ Compute magnitudes

$$||A|| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

For Inception

$$||A|| = \sqrt{0^2 + 0^2 + \dots}$$

$$\Rightarrow \sqrt{7} \approx 2.645$$

For Interstellar:

$$||B|| = \sqrt{7} \approx 2.645 \approx 2.645$$

For Dark Knight

$$||C|| = \sqrt{7} \approx 2.645$$

↳ Compute cosine similarity

$$\text{similarity (Inception, Interstellar)} = \frac{4}{(2.645 \times 2.645)} \Rightarrow 0.57$$

$$\text{similarity (Inception, The Dark Knight)} = \frac{3}{7} \approx 0.43$$

Since, Inception and Interstellar have high similarity (0.57), Interstellar is recommended when searching for movies