

How can I calculate the cosine similarity in R?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I calculate the cosine similarity in R?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137562>

Cosine similarity is a common measure used to determine the similarity between two vectors in data analysis. It is particularly useful in natural language processing and information retrieval. In R, the cosine similarity can be calculated using the "cosine" function from the "proxy" package. This function takes two vectors as input and returns a value between 0 and 1, where 1 indicates a perfect match and 0 indicates no similarity. The calculated cosine similarity can be used to compare text documents, image features, or any other numerical data in R. By understanding and utilizing the cosine similarity calculation in R, users can effectively analyze and compare large datasets for various purposes.

Calculate Cosine Similarity in R

Cosine Similarity is a measure of the similarity between two vectors of an inner product space.

For two vectors, A and B, the Cosine Similarity is calculated as:

$$\text{Cosine Similarity} = \frac{\sum A_i B_i}{\sqrt{\sum A_i^2} \sqrt{\sum B_i^2}}$$

This tutorial explains how to calculate the Cosine Similarity between vectors in R using the cosine() function from the lsa library.

Cosine Similarity Between Two Vectors in R

The following code shows how to calculate the Cosine Similarity between two vectors in R:

```
library(lsa)#define vectors  
a <- c(23, 34, 44, 45, 42, 27, 33, 34)
```

```
b <- c(17, 18, 22, 26, 26, 29, 31, 30)
```

```
#calculate Cosine Similarity
```

```
cosine(a, b)
```

```
0.965195
```

The Cosine Similarity between the two vectors turns out to be 0.965195.

Cosine Similarity of a Matrix in R

The following code shows how to calculate the Cosine Similarity between a matrix of vectors:

```
library(lsa)
```

```
#define matrix
```

```
a <- c(23, 34, 44, 45, 42, 27, 33, 34)
```

```
b <- c(17, 18, 22, 26, 26, 29, 31, 30)
```

```
c <- c(34, 35, 35, 36, 51, 29, 30, 31)
```

```
data <- cbind(a, b, c)
```

```
#calculate Cosine Similarity
```

```
cosine(data)
```

```
a b c
```

a 1.0000000 0.9651950 0.9812406

b 0.9651950 1.0000000 0.9573478

c 0.9812406 0.9573478 1.0000000

Here is how to interpret the output:

The Cosine Similarity between vectors *a* and *b* is 0.9651950. The Cosine Similarity between vectors *a* and *c* is 0.9812406. The Cosine Similarity between vectors *b* and *c* is 0.9573478.

Notes

- 1. The `cosine()` function will work with a square matrix of any size.**
- 2. The `cosine()` function will work on a matrix, but *not* on a data frame. However, you can easily convert a data frame to a matrix in R by using the `as.matrix()` function.**
- 3. Refer to [this Wikipedia page](#) to learn more details about Cosine Similarity.**