Building Features Using Normalization

Janani Ravi
CO-FOUNDER, LOONYCORN
www.loonycorn.com

Overview

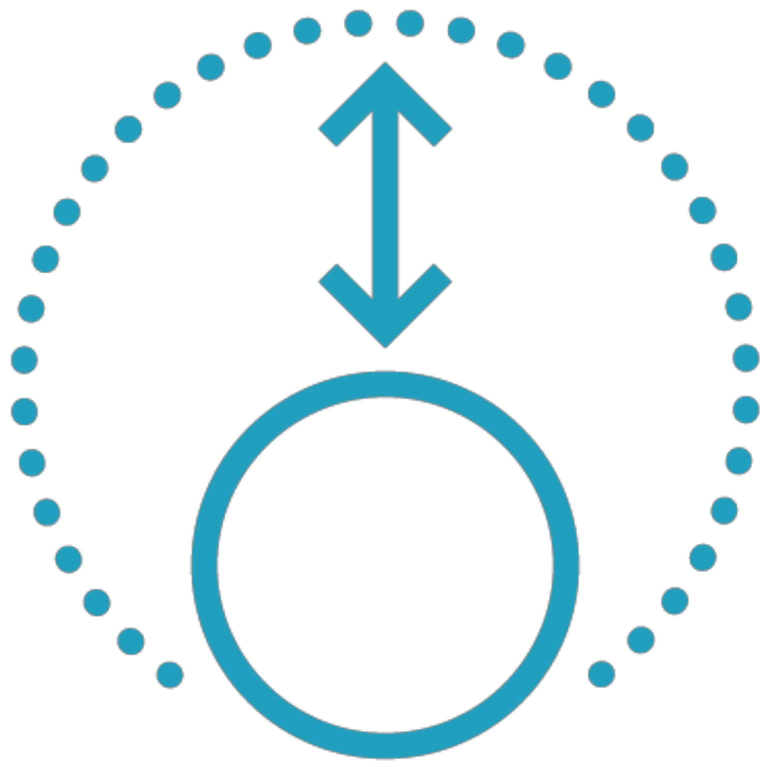
Normalization of feature vectors

Normalization and cosine similarity

L1, L2 and max norms for normalization

What is Normalization?

What is Normalization?

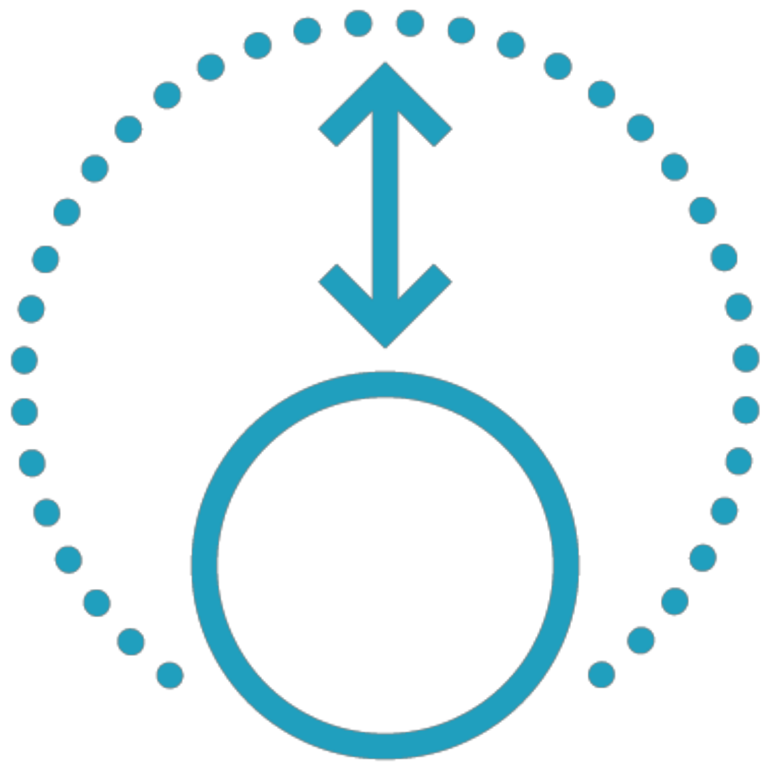


Scaling to a certain range - feature scaling

Centering at 0 and scaling to unit variance - standardization

Transforming a vector to unit norm

What is Normalization?



Scaling to a certain range - feature scaling

Centering at 0 and scaling to unit variance - standardization

Transforming a vector to unit norm

Norm refers to the **magnitude** of the vector

Normalization and Cosine Similarity

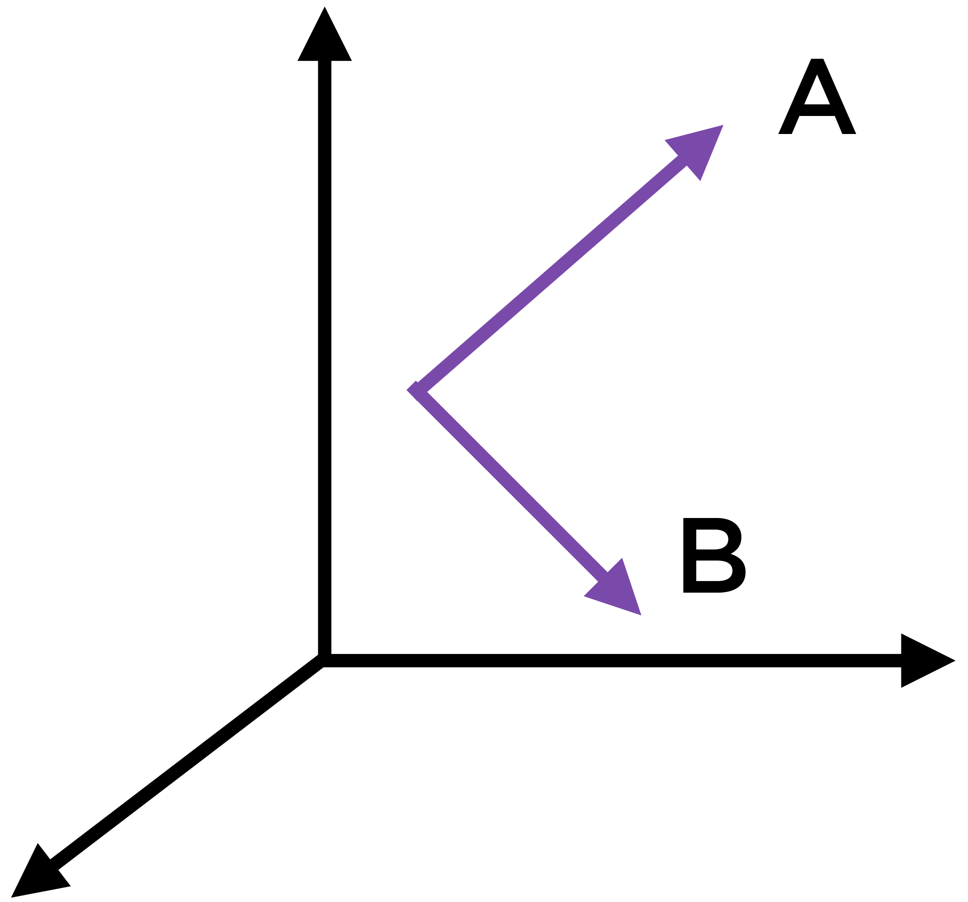
Normalization

Process of scaling input vectors individually to unit norm (unit magnitude), often in order to simplify cosine similarity calculations.

Cosine Similarity

Cosine similarity is a measure of similarity between two non-zero vectors, widely used in ML algorithms - especially in document modeling applications.

Orthogonal Vectors



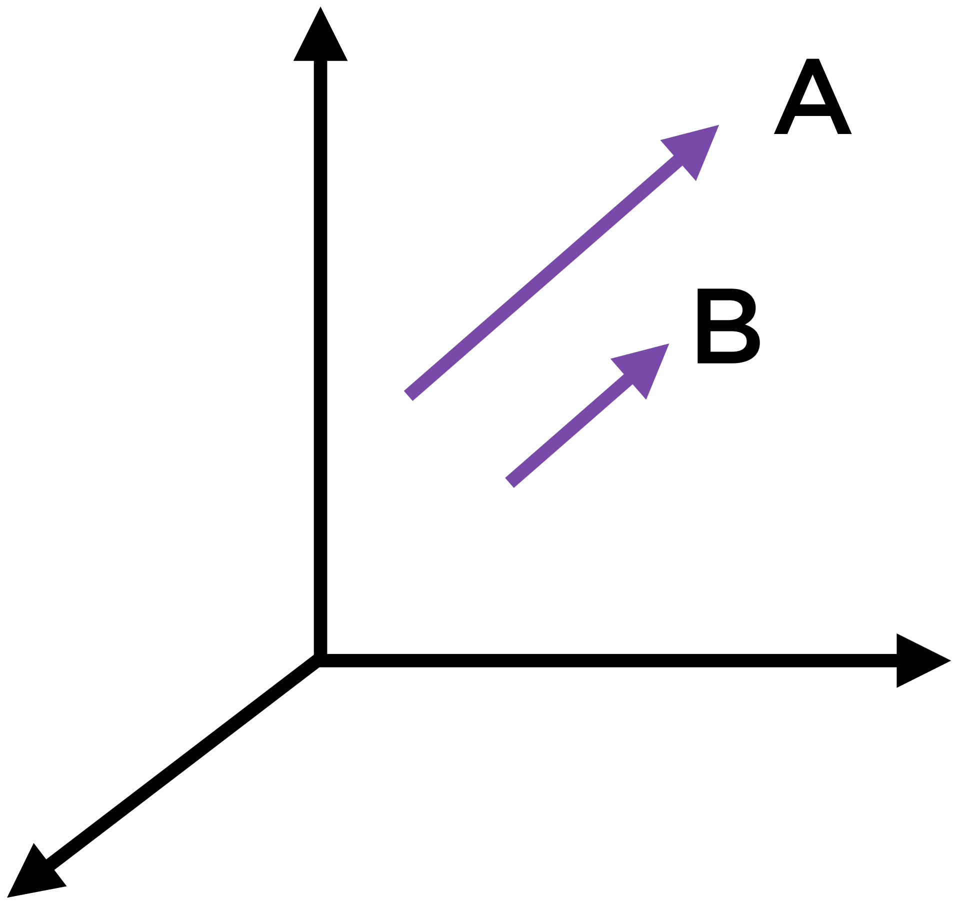
Vectors A and B are at 90 degrees

Orthogonal vectors represent uncorrelated data

A and B are unrelated, independent

Cosine of 90 degrees = 0

Aligned Vectors



Vectors A and B are parallel

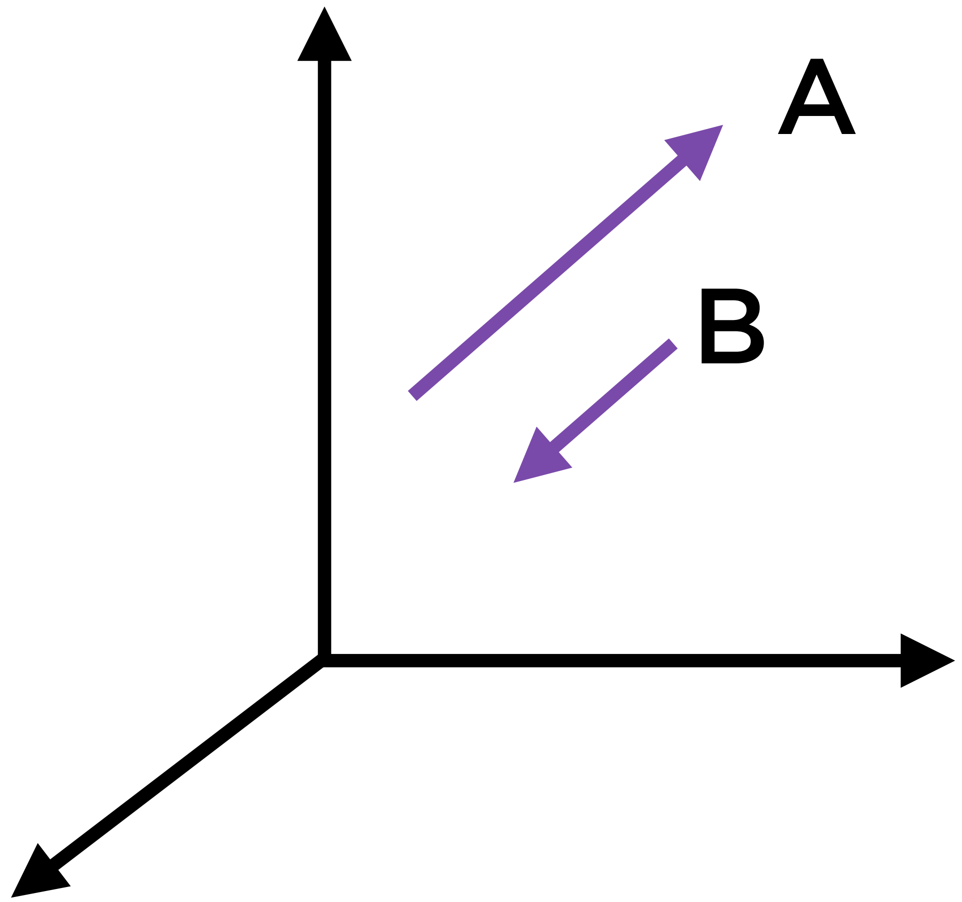
Angle between them is 0 degrees

Perfectly aligned

Correlation of 1 (highest possible)

Cosine of 0 degrees = 1

Opposite Vectors



Vectors A and B point in opposite directions

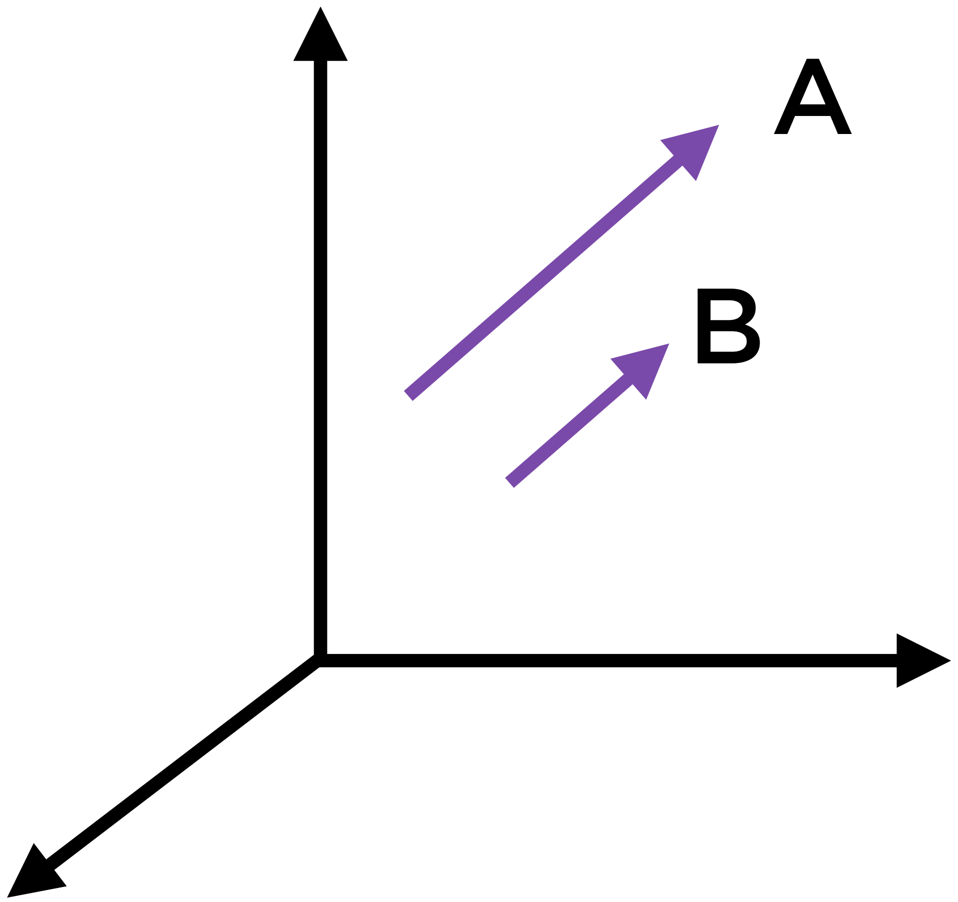
Angle between them is 180 degrees

Perfectly opposed

Correlation of -1 (lowest possible)

Cosine of 180 degrees = -1

Cosine Similarity



Quick and intuitive way to express alignment between two vectors

Each vector represents a single point

In three dimensions, a point is represented as

$$(x_i , y_i , z_i)$$

Cosine Similarity

$$\cos(\theta) = \frac{A \,.\, B}{||A||\ ||B||}$$

$$||A|| = \mathrm{sqrt}(x_A{}^2 + y_A{}^2 + z_A{}^2)$$

$$||B|| = \mathrm{sqrt}(x_B{}^2 + y_B{}^2 + z_B{}^2)$$

$$A.B = x_A x_B + y_A y_B + z_A z_B$$

Cosine Similarity

$$\cos(\theta) = \frac{A \,.\, B}{||A||\ ||B||}$$

$$||A|| = \text{sqrt}(x_A^2 + y_A^2 + z_A^2)$$

$$||B|| = \text{sqrt}(x_B^2 + y_B^2 + z_B^2)$$

$$A.B = x_A x_B + y_A y_B + z_A z_B$$

Simplifying this calculation can simplify the computation of cosine similarity

Normalization

Pre-convert A and B to unit norm vectors to simplify calculation

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{sqrt(x_A^2 + y_A^2 + z_A^2)}$$

$$b = \frac{B}{||B||} = \frac{(x_B, y_B, z_B)}{sqrt(x_B^2 + y_B^2 + z_B^2)}$$

Normalization

$$(x_A, y_A, z_A)$$

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{sqrt(x_A^2 + y_A^2 + z_A^2)}$$

Normalization

$$(x_A, y_A, z_A)$$

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{sqrt(x_A{}^2 + y_A{}^2 + z_A{}^2)}$$

Normalization

$$(x_A, y_A, z_A)$$

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{\mathrm{sqrt}(x_A^2 + y_A^2 + z_A^2)}$$

Normalization

$$(x_A, y_A, z_A)$$

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{sqrt(x_A^2 + y_A^2 + z_A^2)}$$

Normalization

$$(x_A, y_A, z_A)$$

$$a = \frac{A}{||A||} = \frac{(x_A, y_A, z_A)}{sqrt(x_A^2 + y_A^2 + z_A^2)}$$

Cosine Similarity

$$\cos(\theta) = \frac{A \,.\, B}{||A||\ ||B||}$$

$$||A|| = \mathrm{sqrt}(x_A{}^2 + y_A{}^2 + z_A{}^2)$$

$$||B|| = \mathrm{sqrt}(x_B{}^2 + y_B{}^2 + z_B{}^2)$$

$$A.B = x_A x_B + y_A y_B + z_A z_B$$

Cosine Similarity

$$\cos(\theta) = \frac{A \,.\, B}{||A||\ ||B||}$$

$$||A|| = 1$$

$$||B|| = 1$$

$$A.B = x_A x_B + y_A y_B + z_A z_B$$

Normalizing is a row-wise operation, while scaling is a column-wise operation

Data

$$\begin{bmatrix} x_{11} & x_{12} & & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ x_{31} & x_{32} & \cdots & x_{3k} \\ \cdots & \cdots & & \cdots \\ x_{n1} & x_{n2} & & x_{nk} \end{bmatrix}$$

All of the numeric values in our dataset

Columns Represent Features

$$\begin{bmatrix} x_{11} & x_{12} & & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ x_{31} & x_{32} & \cdots & x_{3k} \\ \cdots & \cdots & & \cdots \\ x_{n1} & x_{n2} & & x_{nk} \end{bmatrix}$$

Standardization and scaling apply to an individual feature

Rows Represent Vectors

$$\begin{bmatrix} x_{11} & x_{12} & & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ x_{31} & x_{32} & \cdots & x_{3k} \\ \cdots & \cdots & & \cdots \\ x_{n1} & x_{n2} & & x_{nk} \end{bmatrix}$$

Normalization applies to vectors i.e. to a row which represents data for a single instance

Demo

Normalization and cosine similarity

Types of Normalization

Different Norms

L1

Sum of absolute values of components of vector

L2

Traditional definition of vector magnitude

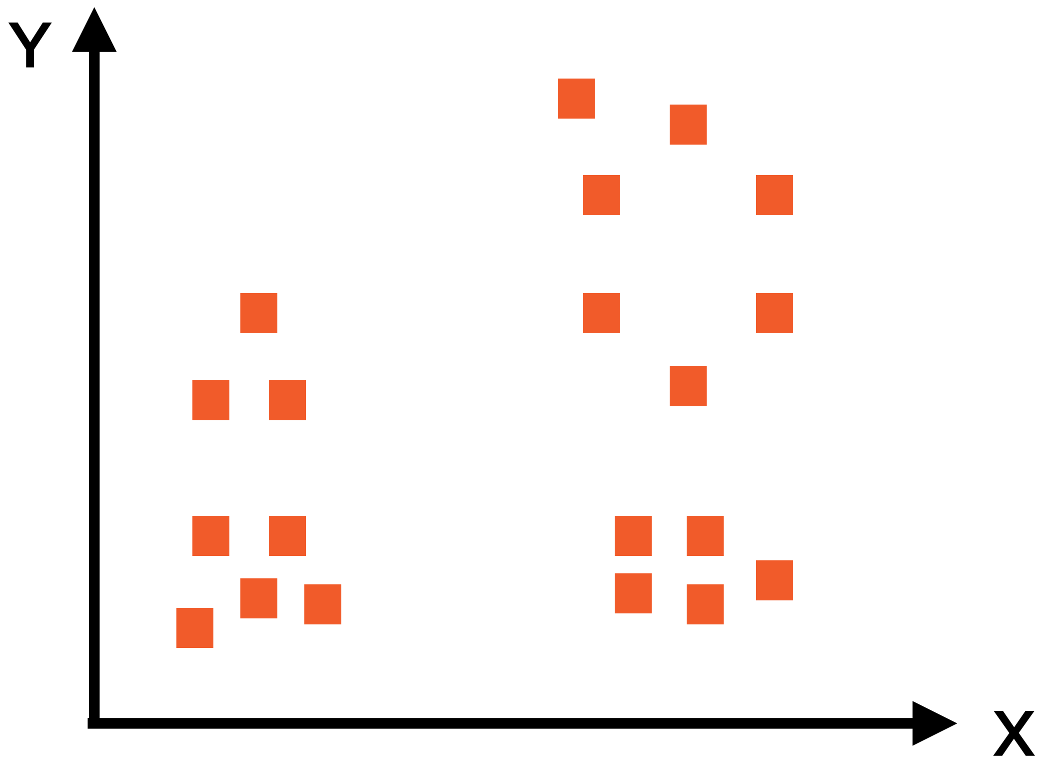
max

Largest absolute value of elements of vector

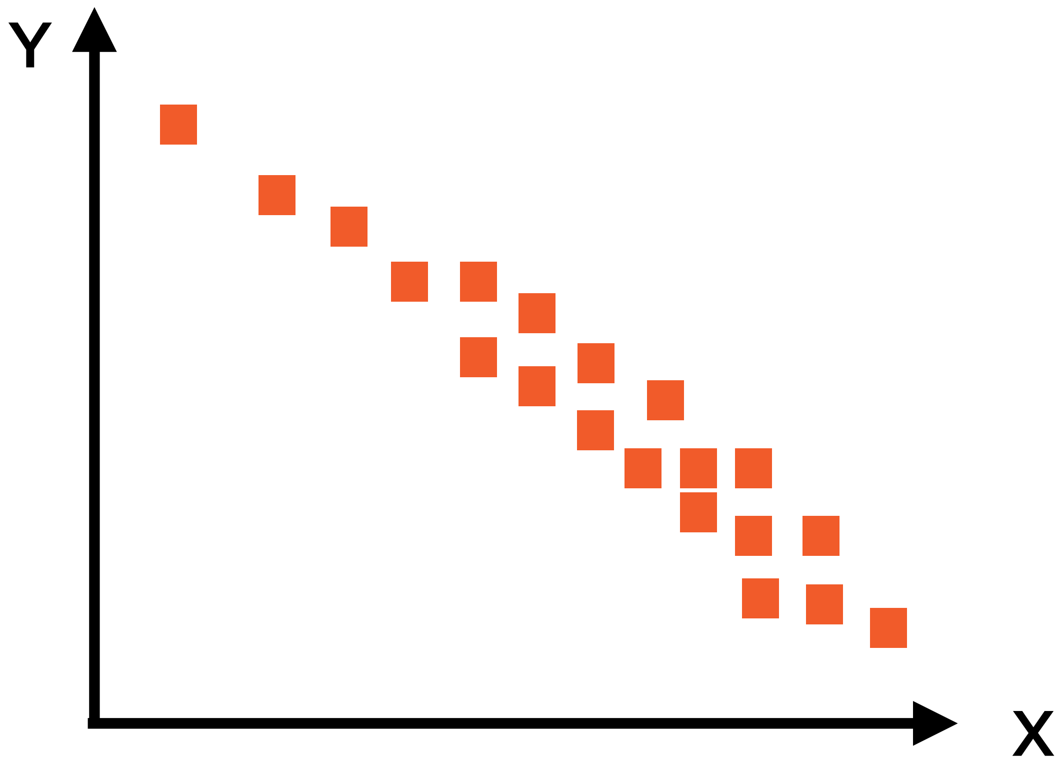
L1-norm

$$x_{new} = \frac{(x, y, z)}{|x| + |y| + |z|}$$

L1-norm



Before L1-norm

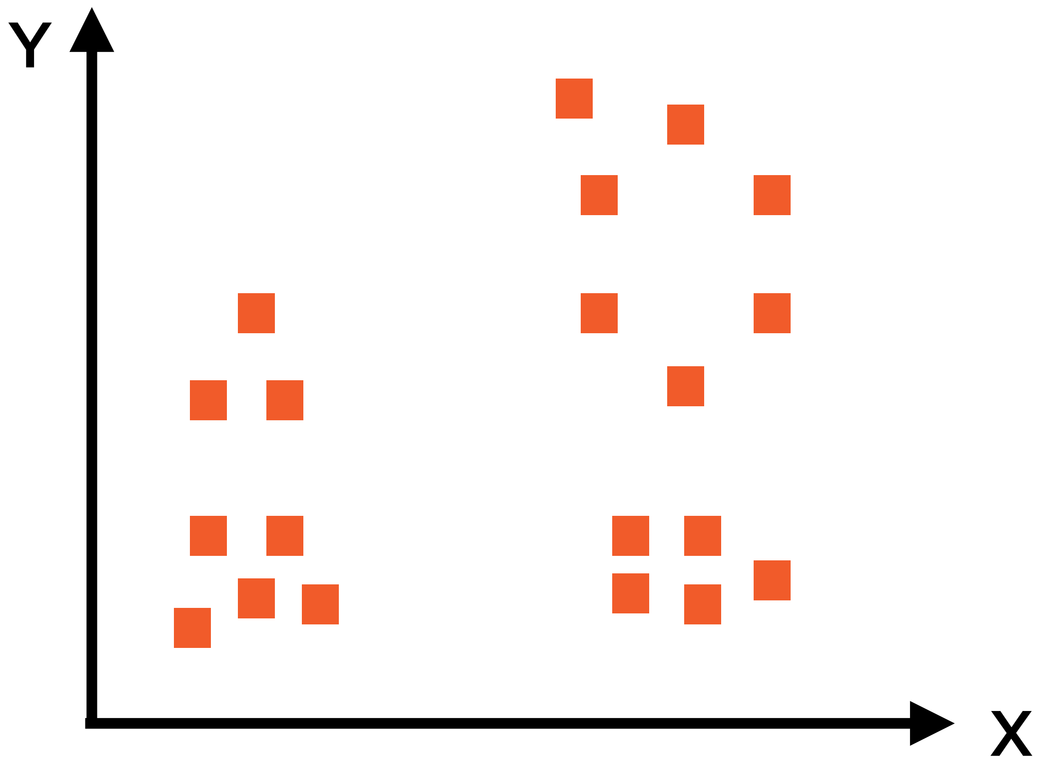


After L1-norm

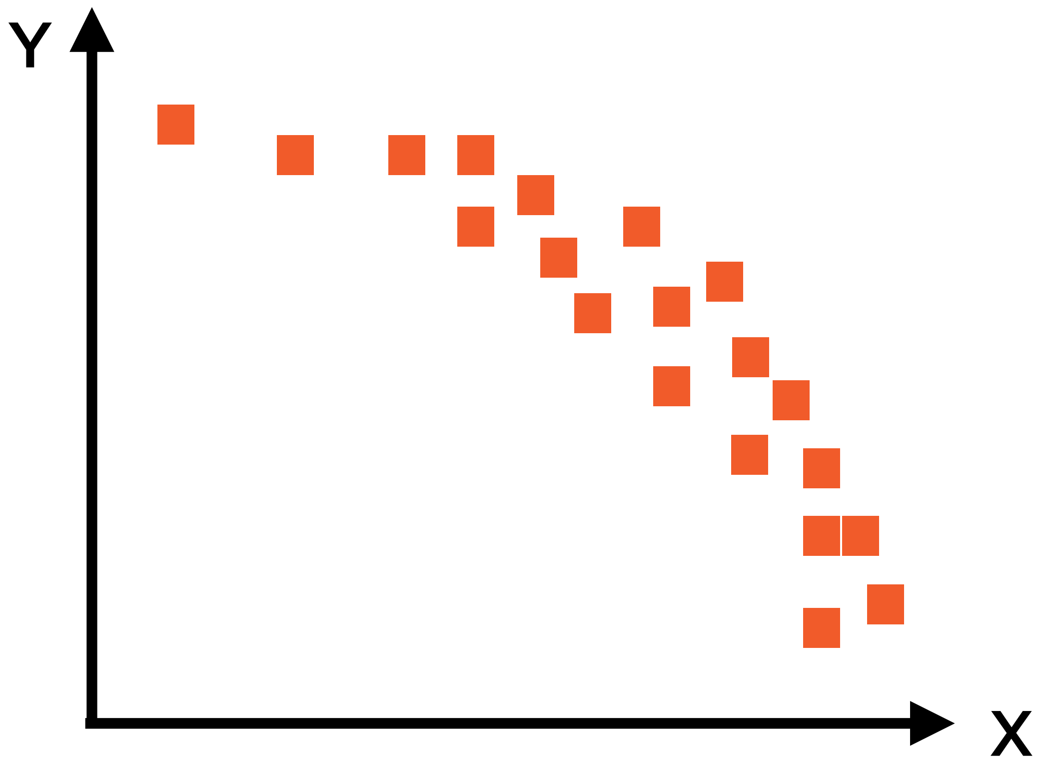
L2-norm

$$x_{new} = \frac{(x, y, z)}{sqrt(x^2 + y^2 + z^2)}$$

L2-norm



Before L2-norm



After L2-norm

max norm

$$x_{new} = \frac{(x, y, z)}{max(abs(x, y, z))}$$

Demo

Applying L1, L2 and max norms for normalization

Summary

Normalization of feature vectors

Normalization and cosine similarity

L1, L2 and max norms for normalization