SAIT

Normalization Techniques

Presented by - Shubh Desai

Types Of Normalization

01 Min Max Normalization

02 L1 Normalization

03 L2 Normalization

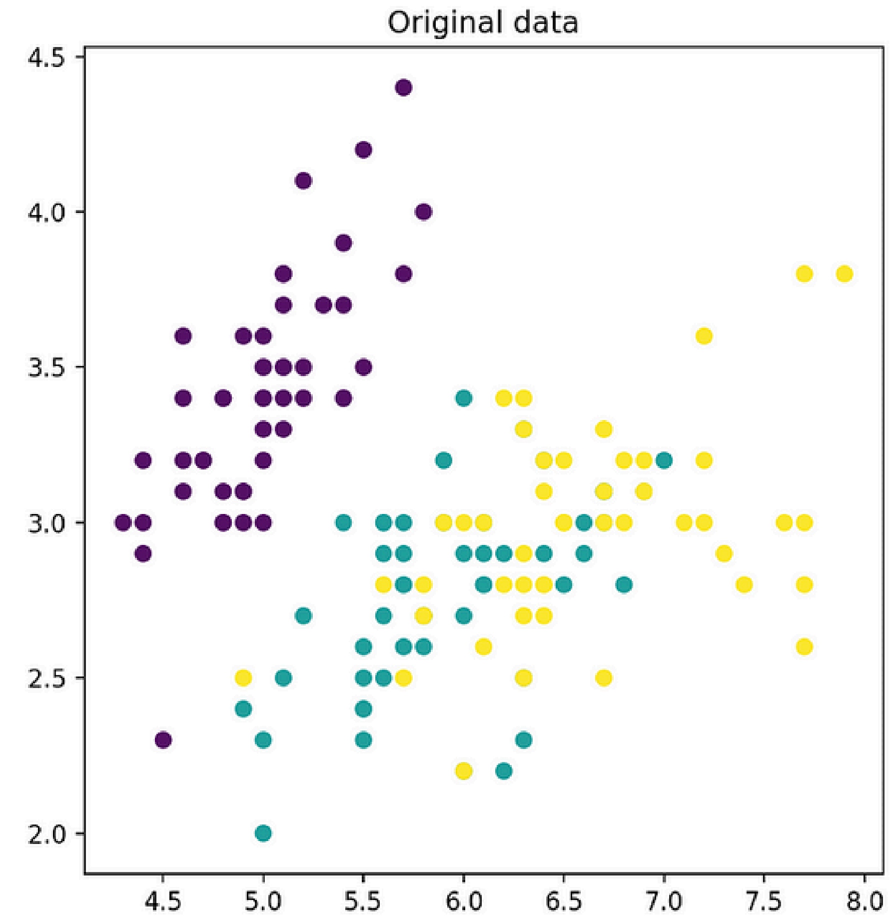
04 Zero Normalization

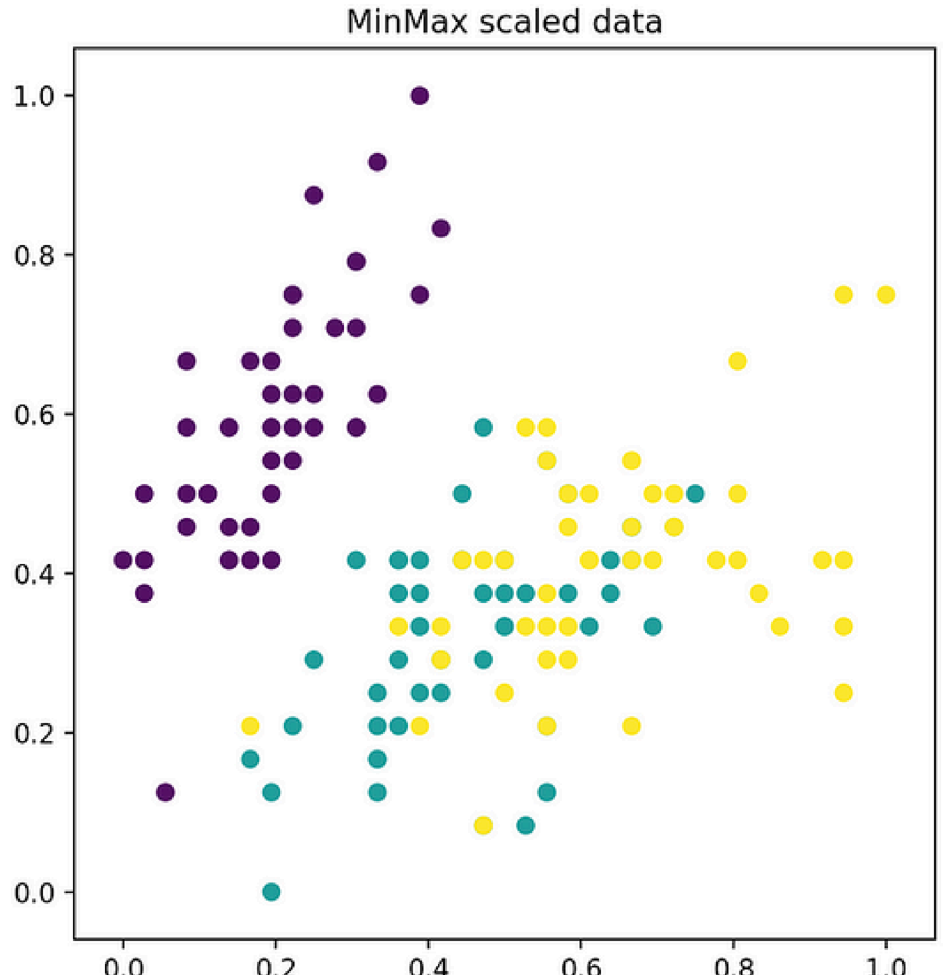
Introduction to Normalization

Definition: Normalization is a technique to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.

Purpose: Helps in faster convergence of gradient descent, removes biases in the dataset, and ensures that all features contribute equally to the model.

Min-Max Normalization





- **Equation Model**: X′= X-Xmin / Xmax - Xmin.
- **Purpose of Model**: Scales the data to a fixed range, usually [0, 1].
- **Uses of Model**: Used in algorithms that require bounded input.
- **Evaluation Metrics**: Mean, Standard Deviation before and after normalization.
- **Examples**: Normalizing features like age and income in a dataset.

Min-Max Normalization Example

```
1000,2000,3000,9000
using min-max normalization by setting min:0 and max:1
```

```
v  =  1000 , putting all values in the formula,we get
v' =  (1000-1000) X (1-0)
     ------------------    + 0  =0
      9000-1000
```

```
v   = 2000, putting all values in the formula,we get
v   '= (2000-1000) X (1-0)
          -----------------         + 0 =0 .125
            9000-1000
```

```
v=3000, putting all values in the formula,we get
v'=(3000-1000) X (1-0)
     -----------------         + 0 =0 .25
       9000-1000
```

```
v=9000, putting all values in the formula, we get
v'=(9000-1000) X (1-0)
  -----------------         + 0 =1
  9000-1000
```

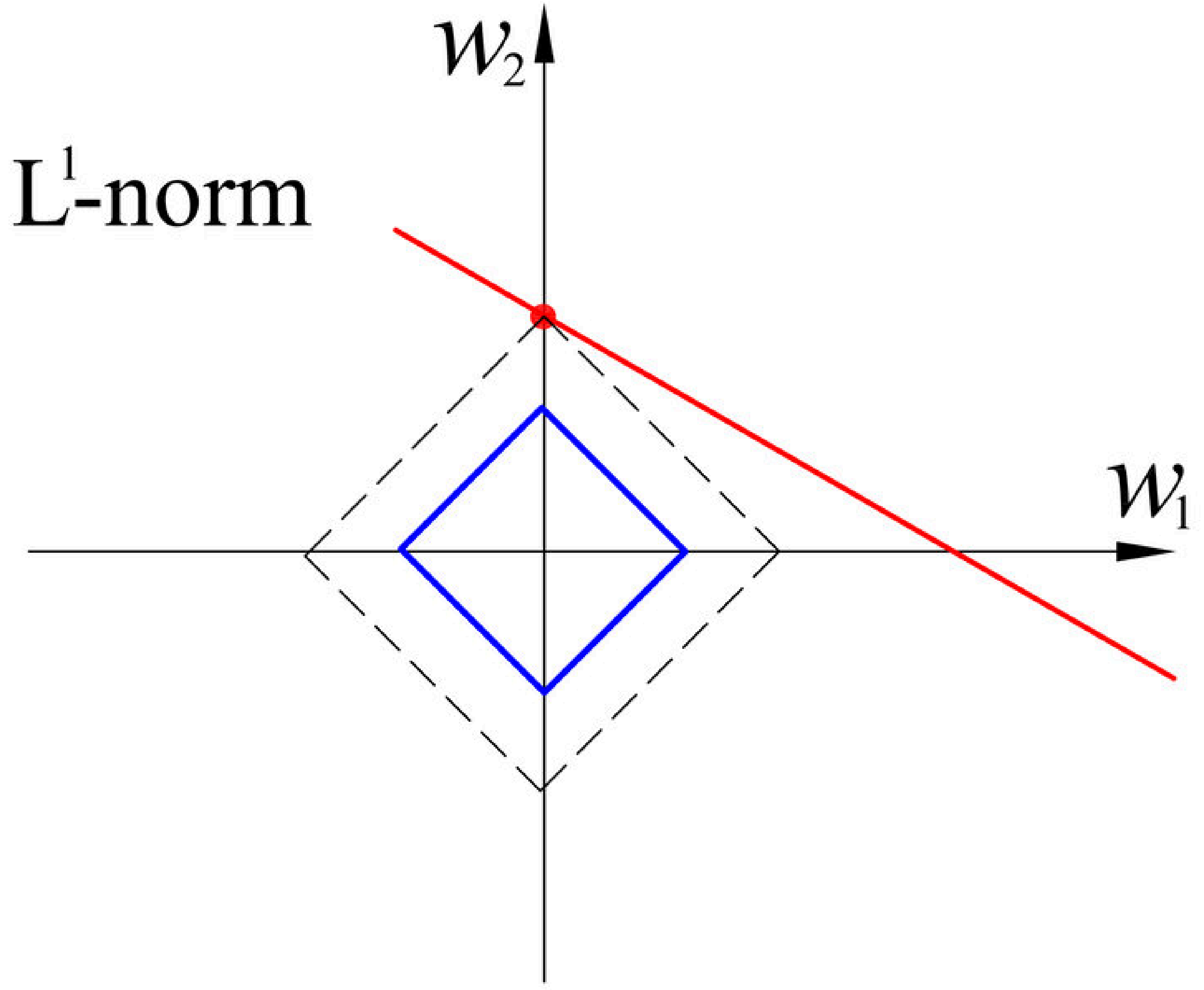
<u>Outcome</u> :

Hence, the normalized values of 1000, 2000, 3000, 9000 are 0, 0.125, .25, 1.

L1 Normalization

- **Purpose of Model**: Makes the sum of the absolute values of all features equal to 1

- **Uses of Model**: L1 normalization is often used in scenarios where the model needs to focus on the most significant features and ignore the rest, such as in text classification



L1 Normalization

1. L1 Normalization Example

Input Vector: $[3, 4]$

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|_1} = \frac{[x_1, x_2, \ldots, x_n]}{|x_1| + |x_2| + \cdots + |x_n|}$$

Steps:

- Calculate the sum of absolute values: $|3| + |4| = 7$.
- Divide each element by this sum:

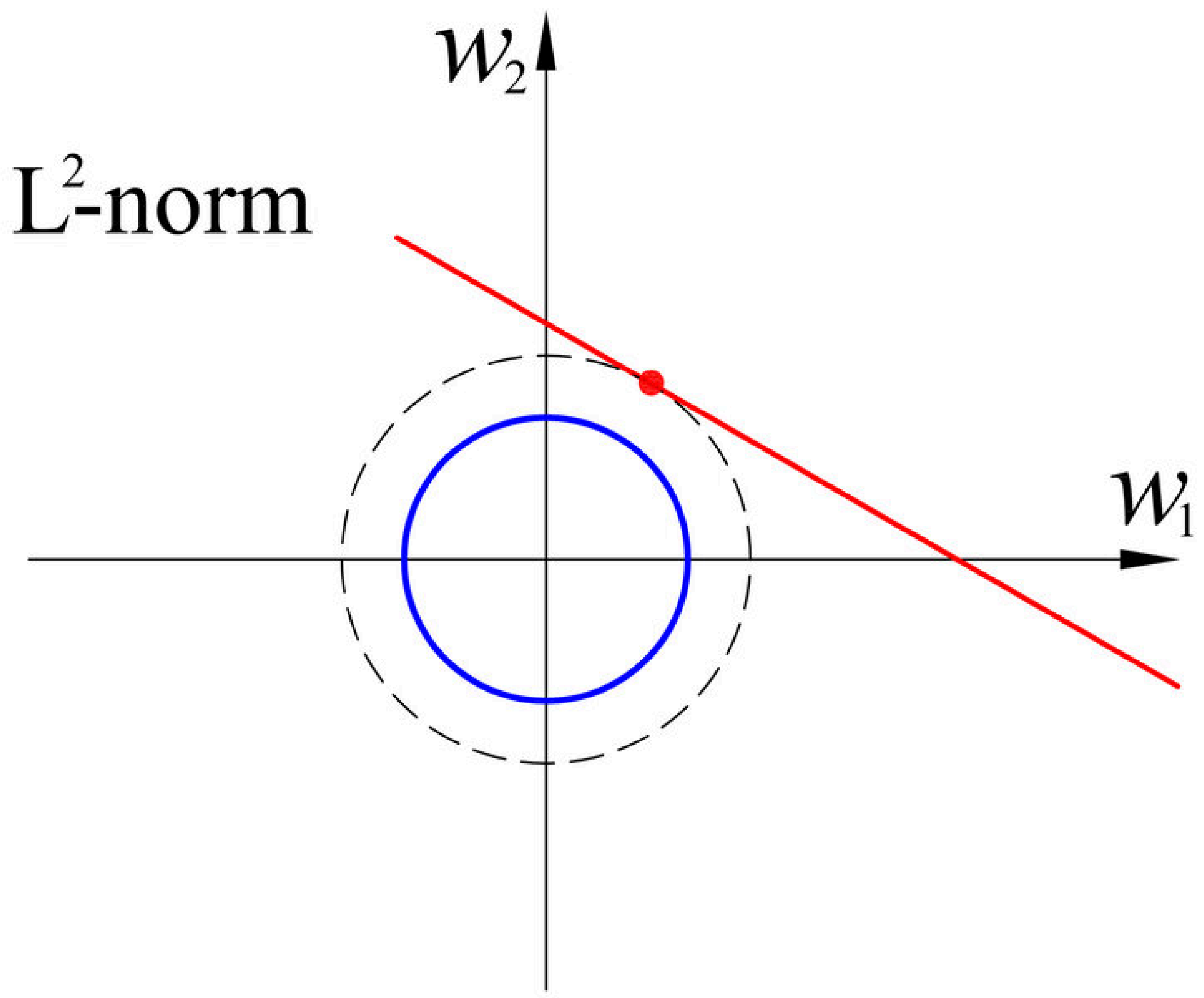
$$\text{Normalized vector} = \left[\frac{3}{7}, \frac{4}{7}\right] = [0.4286, 0.5714]$$

Outcome: The vector is scaled so that the absolute sum of all elements equals 1.

L2 Normalization

- **Purpose of Model**: Makes the sum of the squares of all features equal to 1

- **Uses of Model**: L2 normalization is commonly used in machine learning algorithms like k-means clustering and k-nearest neighbors where you care about the distance between data points, like in clustering.



L2 Normalization

2. L2 Normalization Example

Input Vector: $[3, 4]$

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} = \frac{[x_1, x_2, \ldots, x_n]}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}}$$

Steps:

- Calculate the square root of the sum of squares: $\sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5.$
- Divide each element by this value:

$$\text{Normalized vector} = \left[\frac{3}{5}, \frac{4}{5}\right] = [0.6, 0.8]$$

Outcome: The vector is scaled so that the sum of the squares of the elements equals 1.

Zero Normalzation (Z-Score)

- **Purpose of Model:** Adjusts data so that the mean of the features is 0

- **Uses of Model:** Used when features have different scales and units, e.g., linear regression, logistic regression.

Zero Normalzation (Z-Score)

3. Zero Normalization Example

Input Vector: $[3, 4, 5]$

$$\mathbf{x}' = \mathbf{x} - \mu = [x_1 - \mu, x_2 - \mu, \ldots, x_n - \mu]$$

Steps:

- Calculate the mean: $\frac{3+4+5}{3} = 4$.
- Subtract the mean from each element:

$$\text{Normalized vector} = [3-4, 4-4, 5-4] = [-1, 0, 1]$$

Outcome: The data is centered around zero, with the mean of the vector becoming 0.

Thank You