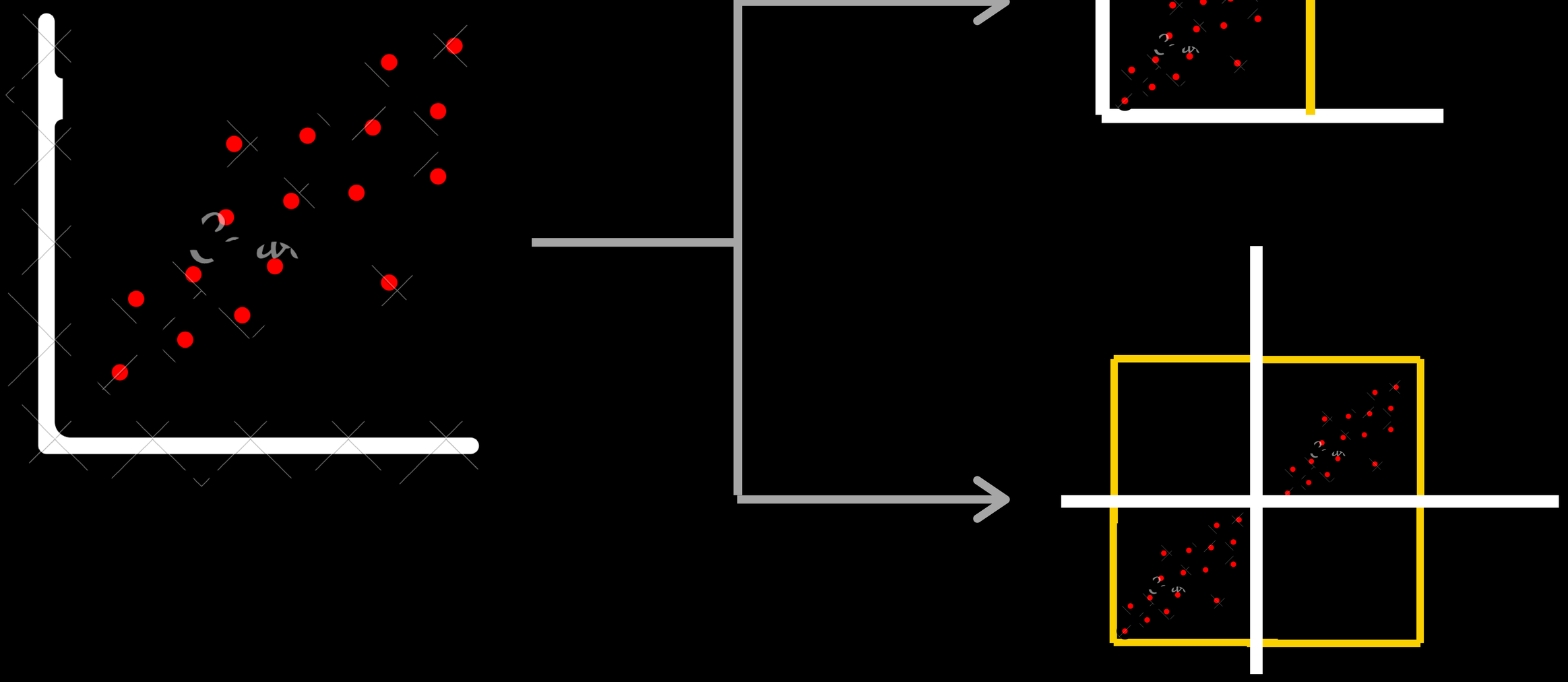
01

5 mins read

FEATURE SCALING IN MACHINE LEARNING

Normalization and standardization



→ Demo

@learn.machinelearning

What is feature scaling?

It is one of the most fundamental steps in the part of data processing.

Goal of this technique is to make sure all features are on almost the same scale.

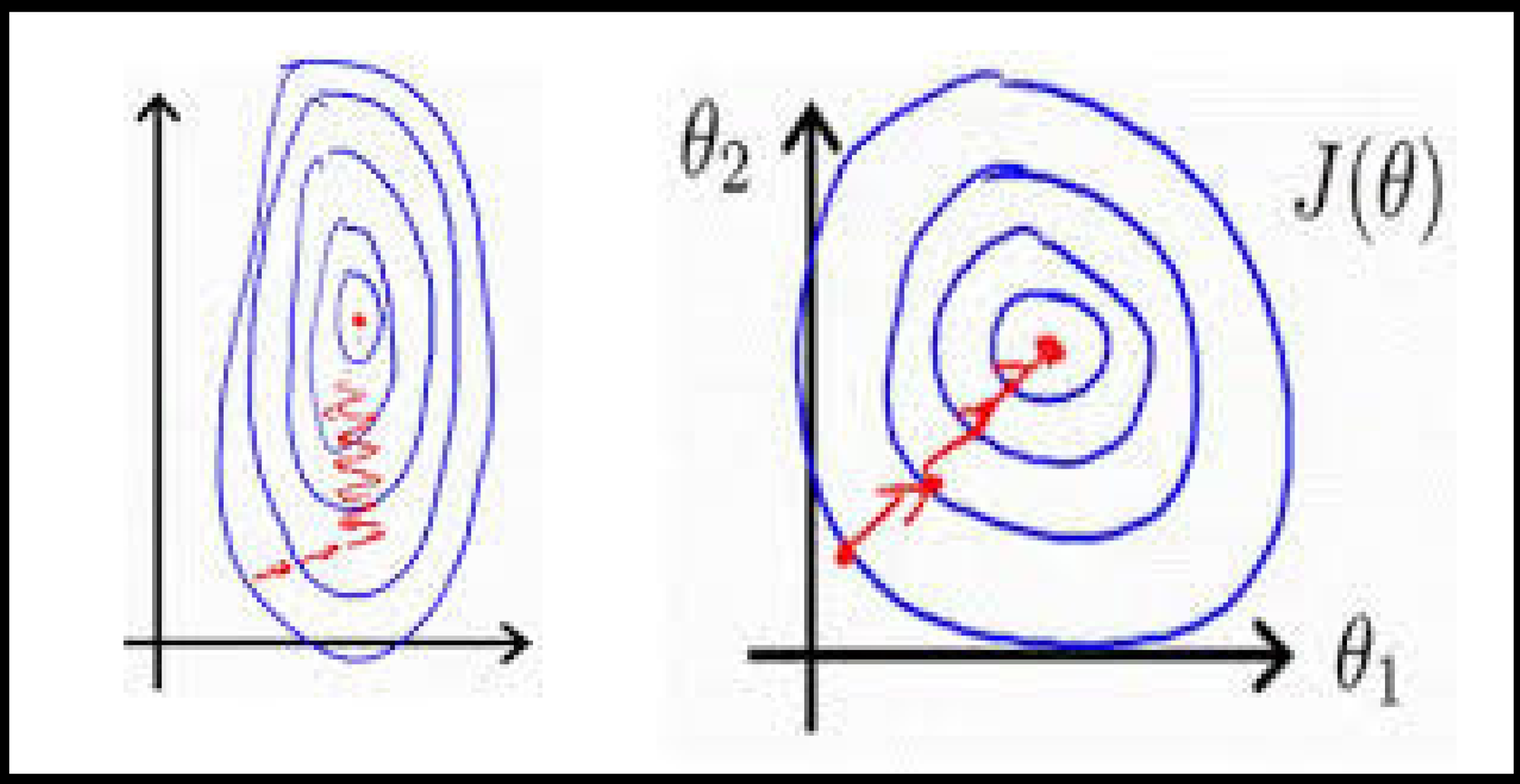
Because most of the times our dataset will contain features highly varying in magnitudes, units and range.

Why feature scaling?

- To give equal importance to each feature.
- This technique also makes it easy for ML algorithms to process the data.
- To solve issues faced by different algorithms.
- Helps algorithms to train and converge faster.

Problems with Gradient Descent Based Algorithms

- Look at the picture
- Different scale features will affect the step size of a gradient descent.
- Scaling features can help the gradient descent converge more quickly towards the minima.
- Example algorithms like linear regression, logistic regression, neural network, etc.



Problems with Distance-Based Algorithms

- These algorithms are biased towards large values.
- Because we calculate the distance between the data points to determine their similarity.
- This will give different weights to different features.
- Example algorithms like KNN, K-means, and SVM

When it is not important?

If an algorithm is not distance-based, feature scaling is unimportant, including Naive Bayes, Linear Discriminant Analysis, and Tree-Based models (gradient boosting, random forest, etc.).

Normalization

- Min-Max scaling.
- Which is used to transform all features to be on a similar scale (they range to [0,1]).
- This is affected by outliers.
- Maximum value will be 1 and the minimum will be 0

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Standardization

- We also call it as Z-Score Normalization.
- Which is used to transform features by subtracting from mean and dividing by standard deviation.
- So, all the values are centred around the mean with a unit SD.
- Each feature will be of different scale.It can handle outliers

$$z = \frac{x_i - \mu}{\sigma}$$

New video is out

Learn more about when to use the normalization and when to use standardization with code demo