

Title: Feature scaling

URL: [https://en.wikipedia.org/wiki/Feature\\_scaling](https://en.wikipedia.org/wiki/Feature_scaling)

PageID: 34061548

Categories: Category:Machine learning, Category:Statistical data transformation

Source: Wikipedia (CC BY-SA 4.0). Content may require attribution.

-----

Supervised learning

Unsupervised learning

Semi-supervised learning

Self-supervised learning

Reinforcement learning

Meta-learning

Online learning

Batch learning

Curriculum learning

Rule-based learning

Neuro-symbolic AI

Neuromorphic engineering

Quantum machine learning

Classification

Generative modeling

Regression

Clustering

Dimensionality reduction

Density estimation

Anomaly detection

Data cleaning

AutoML

Association rules

Semantic analysis

Structured prediction

Feature engineering

Feature learning

Learning to rank

Grammar induction

Ontology learning

Multimodal learning

Apprenticeship learning

Decision trees

Ensembles Bagging Boosting Random forest

Bagging

Boosting

Random forest

k -NN

Linear regression

Naive Bayes

Artificial neural networks

Logistic regression

Perceptron

Relevance vector machine (RVM)

Support vector machine (SVM)

BIRCH

CURE

Hierarchical

k -means

Fuzzy

Expectation–maximization (EM)

DBSCAN

OPTICS

Mean shift

Factor analysis

CCA

ICA

LDA

NMF

PCA

PGD

t-SNE

SDL

Graphical models Bayes net Conditional random field Hidden Markov

Bayes net

Conditional random field

Hidden Markov

RANSAC

k -NN

Local outlier factor  
Isolation forest  
Autoencoder  
Deep learning  
Feedforward neural network  
Recurrent neural network LSTM GRU ESN reservoir computing  
LSTM  
GRU  
ESN  
reservoir computing  
Boltzmann machine Restricted  
Restricted  
GAN  
Diffusion model  
SOM  
Convolutional neural network U-Net LeNet AlexNet DeepDream  
U-Net  
LeNet  
AlexNet  
DeepDream  
Neural field Neural radiance field Physics-informed neural networks  
Neural radiance field  
Physics-informed neural networks  
Transformer Vision  
Vision  
Mamba  
Spiking neural network  
Memtransistor  
Electrochemical RAM (ECRAM)  
Q-learning  
Policy gradient  
SARSA  
Temporal difference (TD)  
Multi-agent Self-play  
Self-play  
Active learning  
Crowdsourcing  
Human-in-the-loop

Mechanistic interpretability

RLHF

Coefficient of determination

Confusion matrix

Learning curve

ROC curve

Kernel machines

Bias–variance tradeoff

Computational learning theory

Empirical risk minimization

Occam learning

PAC learning

Statistical learning

VC theory

Topological deep learning

AAAI

ECML PKDD

NeurIPS

ICML

ICLR

IJCAI

ML

JMLR

Glossary of artificial intelligence

List of datasets for machine-learning research List of datasets in computer vision and image processing

List of datasets in computer vision and image processing

Outline of machine learning

v

t

e

Feature scaling is a method used to normalize the range of independent variables or features of data. In data processing , it is also known as data normalization and is generally performed during the data preprocessing step.

Motivation

Since the range of values of raw data varies widely, in some machine learning algorithms, objective functions will not work properly without normalization . For example, many classifiers calculate the distance between two points by the Euclidean distance . If one of the features has a broad range of values, the distance will be governed by this particular feature. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance.

Another reason why feature scaling is applied is that gradient descent converges much faster with feature scaling than without it.

It's also important to apply feature scaling if regularization is used as part of the loss function (so that coefficients are penalized appropriately).

Empirically, feature scaling can improve the convergence speed of stochastic gradient descent. In support vector machines, it can reduce the time to find support vectors. Feature scaling is also often used in applications involving distances and similarities between data points, such as clustering and similarity search. As an example, the K-means clustering algorithm is sensitive to feature scales.

## Methods

### Rescaling (min-max normalization)

Also known as min-max scaling or min-max normalization, rescaling is the simplest method and consists in rescaling the range of features to scale the range in  $[0, 1]$  or  $[-1, 1]$ . Selecting the target range depends on the nature of the data. The general formula for a min-max of  $[0, 1]$  is given as:

where  $x$  is an original value,  $x'$  is the normalized value. For example, suppose that we have the students' weight data, and the students' weights span  $[160 \text{ pounds}, 200 \text{ pounds}]$ . To rescale this data, we first subtract 160 from each student's weight and divide the result by 40 (the difference between the maximum and minimum weights).

To rescale a range between an arbitrary set of values  $[a, b]$ , the formula becomes:

where  $a, b$  are the min-max values.

### Mean normalization

where  $x$  is an original value,  $x'$  is the normalized value,  $\bar{x} = \text{average}(x)$  is the mean of that feature vector. There is another form of the means normalization which divides by the standard deviation which is also called standardization.

### Standardization (Z-score Normalization)

In machine learning, we can handle various types of data, e.g. audio signals and pixel values for image data, and this data can include multiple dimensions. Feature standardization makes the values of each feature in the data have zero-mean (when subtracting the mean in the numerator) and unit-variance. This method is widely used for normalization in many machine learning algorithms (e.g., support vector machines, logistic regression, and artificial neural networks). The general method of calculation is to determine the distribution mean and standard deviation for each feature. Next we subtract the mean from each feature. Then we divide the values (mean is already subtracted) of each feature by its standard deviation.

Where  $x$  is the original feature vector,  $\bar{x} = \text{average}(x)$  is the mean of that feature vector, and  $\sigma$  is its standard deviation.

### Robust Scaling

Robust scaling, also known as standardization using median and interquartile range (IQR), is designed to be robust to outliers. It scales features using the median and IQR as reference points instead of the mean and standard deviation:  $x' = \frac{x - Q_2(x)}{Q_3(x) - Q_1(x)}$  where  $Q_1(x)$ ,  $Q_2(x)$ ,  $Q_3(x)$  are the three quartiles (25th, 50th, 75th percentile) of the feature.

### Unit vector normalization

Unit vector normalization regards each individual data point as a vector, and divide each by its vector norm, to obtain  $x' = x / \|x\|$ . Any vector norm can be used, but the most common ones are the L1 norm and the L2 norm.

For example, if  $x = (v_1, v_2, v_3)$ , then its  $L_p$ -normalized version is:  $\left(\frac{v_1}{|v_1|^p + |v_2|^p + |v_3|^p}^{1/p}, \frac{v_2}{|v_1|^p + |v_2|^p + |v_3|^p}^{1/p}, \frac{v_3}{|v_1|^p + |v_2|^p + |v_3|^p}^{1/p}\right)$

See also

Normalization (machine learning)

Normalization (statistics)

Standard score

fMLLR , Feature space Maximum Likelihood Linear Regression

References

Further reading

Han, Jiawei; Kamber, Micheline; Pei, Jian (2011). "Data Transformation and Data Discretization" . Data Mining: Concepts and Techniques . Elsevier. pp. 111– 118. ISBN 9780123814807 .

External links

Lecture by Andrew Ng on feature scaling Archived 2017-03-14 at the Wayback Machine