

Feature scaling

Feature scaling is a method used to standardize 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.

Contents

Motivation

Methods

- Rescaling (min-max normalization)

- Mean normalization

- Standardization

- Scaling to unit length

Application

See also

References

General references

External links

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, the majority of [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.^[1]

Methods

Rescaling (min-max normalization)

Also known as min-max scaling or min-max normalization, 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 is given as:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

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 pounds, 200 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).

Mean normalization

$$x' = \frac{x - \text{average}(x)}{\max(x) - \min(x)}$$

where x is an original value, x' is the normalized value.

Standardization

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)^[2]. 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.

$$x' = \frac{x - \bar{x}}{\sigma}$$

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

Scaling to unit length

Another option that is widely used in machine-learning is to scale the components of a feature vector such that the complete vector has length one. This usually means dividing each component by the Euclidean length of the vector:

$$x' = \frac{x}{||x||}$$

In some applications (e.g. Histogram features) it can be more practical to use the L1 norm (i.e. Manhattan Distance, City-Block Length or Taxicab Geometry) of the feature vector. This is especially important if in the following learning steps the Scalar Metric is used as a distance measure.

Application

In stochastic gradient descent, feature scaling can sometimes improve the convergence speed of the algorithm^[2]. In support vector machines,^[3] it can reduce the time to find support vectors. Note that feature scaling changes the SVM result

See also

- fMLLR, Feature space Maximum Likelihood Linear Regression

References

1. Ioffe, Sergey; Christian Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". arXiv:1502.03167 (<https://arxiv.org/abs/1502.03167>) [cs.LG (<https://arxiv.org/archive/cs.LG>)].
2. Grus, Joel (2015). *Data Science from Scratch* Sebastopol, CA: O'Reilly pp. 99, 100. ISBN 978-1-491-90142-7.
3. Juszczak, P.; D. M. J. Tax; R. P. W. Dui (2002). "Feature scaling in support vector data descriptions". *Proc. 8th Annu. Conf. Adv. School Comput. Imaging* 25–30. CiteSeerX 10.1.1.100.2524 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.100.2524>)

General references

- S. Aksoy and R. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognit. Lett.*, Special Issue on Image and Video Retrieval, 2000
http://www.cs.bilkent.edu.tr/~saksoy/papers/pletters01_likelihood.pdf
- S. Tsakalidis, V. Doumptotis & W. Byrne, "Discriminative Linear Transforms for Feature Normalization and Speaker Adaptation in HMM Estimation", *Proc. ICSLP'02*, Denver
http://malach.umiacs.umd.edu/pubs/VD_05_Discrim_linea.pdf
- Liefeng Bo, Ling Wang, and Licheng Jiao, "Feature Scaling for Kernel Fisher Discriminant Analysis Using Leave-one-out Cross Validation", *Neural Computation (NECO)*, vol. 18(4), pp. 961–978, 2006
<http://www.cs.washington.edu/homes/lfb/paper/nc06.pdf>
- A. Stolcke, S. Kajarekar and L. Ferrer, "Nonparametric feature normalization for SVM-based speaker verification," in *Proc. ICASSP*, Las Vegas, Apr. 2008. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4517925
- Youn, E.; Jeong, M. K. (2009). "Class dependent feature scaling method using naive Bayes classifier for text datamining". *Pattern Recognition Letters* **30** (5): 477–485. doi:10.1016/j.patrec.2008.11.013
- S. Theodoridis, K. Koutroubas. (2008) "Pattern Recognition", Academic Press, 4 edition, ISBN 978-1-59749-272-0

External links

- [Lecture by Andrew Ng on feature scaling](#)
-

Retrieved from 'https://en.wikipedia.org/w/index.php?title=Feature_scaling&oldid=870599160

This page was last edited on 25 November 2018, at 21:46 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.