



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export

Create a book
Download as PDF
Printable version

Languages

Deutsch
Español
한국어
日本語

Edit links

Create account Log in

Article **Talk**

Read **Edit** View history

Search

Bootstrap aggregating

From Wikipedia, the free encyclopedia

Bootstrap aggregating, also called **bagging**, is a [machine learning ensemble meta-algorithm](#) designed to improve the stability and accuracy of [machine learning](#) algorithms used in [statistical classification](#) and [regression](#). It also reduces [variance](#) and helps to avoid [overfitting](#). Although it is usually applied to [decision tree](#) methods, it can be used with any type of method. Bagging is a special case of the [model averaging](#) approach.

Contents [hide]

- Description of the technique
- Example: Ozone data
- Bagging for nearest neighbour classifiers
- History
- See also
- References

Description of the technique edit

Given a standard [training set](#) D of size n , bagging generates m new training sets D_i , each of size n' , by [sampling](#) from D [uniformly](#) and [with replacement](#). By sampling with replacement, some observations may be repeated in each D_i . If $n'=n$, then for large n the set D_i is expected to have the fraction $(1 - 1/e)$ ($\approx 63.2\%$) of the unique examples of D , the rest being duplicates.^[1] This kind of sample is known as a [bootstrap](#) sample. The m models are fitted using the above m bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

Bagging leads to "improvements for unstable procedures" (Breiman, 1996), which include, for example, [artificial neural networks](#), [classification and regression trees](#), and subset selection in [linear regression](#) (Breiman, 1994). An interesting application of bagging showing improvement in preimage learning is provided here.^{[2][3]} On the other hand, it can mildly degrade the performance of stable methods such as K-nearest neighbors (Breiman, 1996).

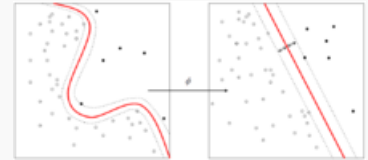
Example: Ozone data edit

To illustrate the basic principles of bagging, below is an analysis on the relationship between [ozone](#) and temperature (data from [Rousseeuw](#) and Leroy (1986), available at [classic data sets](#), analysis done in [R](#)).

The relationship between temperature and ozone in this data set is apparently non-linear, based on the scatter plot. To mathematically describe this relationship, [LOESS](#) smoothers (with span 0.5) are used. Instead of building a single smoother from the complete data set, 100 [bootstrap](#) samples of the data were drawn. Each sample is different from the original data set, yet resembles it in distribution and variability. For each bootstrap sample, a LOESS smoother was fit. Predictions from these 100 smoothers were then made across the range of the data. The first 10 predicted smooth fits appear as grey lines in the figure below. The lines are clearly very *wiggly* and they overfit the data - a result of the span being too low.

By taking the average of 100 smoothers, each fitted to a subset of the original data set, we arrive at one

Machine learning and data mining



Problems

Classification · Clustering · Regression · Anomaly detection · Association rules · Reinforcement learning · Structured prediction · Feature learning · Online learning · Semi-supervised learning · Unsupervised learning · Learning to rank · Grammar induction

Supervised learning

(classification · regression)

Decision trees · Ensembles (**Bagging**, Boosting, Random forest) · k -NN · Linear regression · Naive Bayes · Neural networks · Logistic regression · Perceptron · Support vector machine (SVM) · Relevance vector machine (RVM)

Clustering

BIRCH · Hierarchical · k -means · Expectation-maximization (EM) · DBSCAN · OPTICS · Mean-shift

Dimensionality reduction

Factor analysis · CCA · ICA · LDA · NMF · PCA · t-SNE

Structured prediction

Graphical models (Bayes net, CRF, HMM)

Anomaly detection

k -NN · Local outlier factor

Neural nets

Autoencoder · Deep learning · Multilayer perceptron · RNN · Restricted Boltzmann machine · SOM · Convolutional neural network

Theory

Bias-variance dilemma · Computational learning theory · Empirical risk minimization · PAC learning · Statistical learning · VC theory



[Machine learning portal](#)



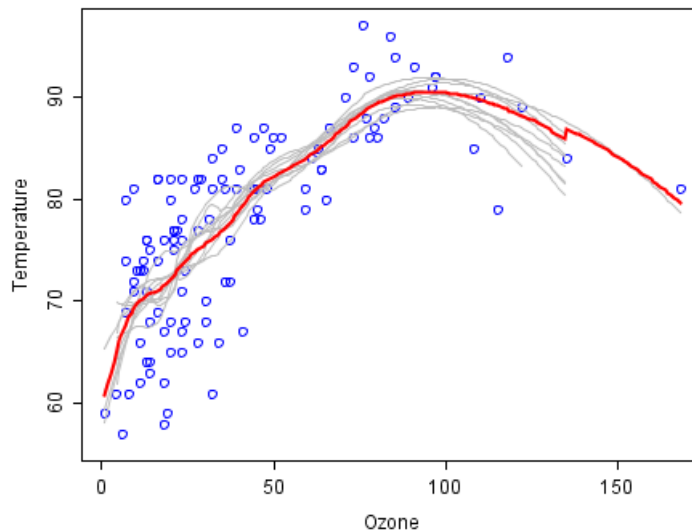
[Computer science portal](#)



[Statistics portal](#)

v · t · e

bagged predictor (red line). Clearly, the mean is more stable and there is less [overfit](#).



Bagging for nearest neighbour classifiers [\[edit\]](#)

The [risk](#) of a 1 nearest neighbour (1NN) classifier is at most twice the risk of the [Bayes classifier](#),^[4] but there are no guarantees that this classifier will be [consistent](#). By careful choice of the size of the resamples, bagging can lead to substantial improvements of the performance of the 1NN classifier. By taking a large number of resamples of the data of size n' , the bagged nearest neighbour classifier will be consistent provided $n' \rightarrow \infty$ diverges but $n'/n \rightarrow 0$ as the sample size $n \rightarrow \infty$.

Under infinite simulation, the bagged nearest neighbour classifier can be viewed as a [weighted nearest neighbour classifier](#). Suppose that the feature space is d dimensional and denote by $C_{n,n'}^{bnn}$ the bagged nearest neighbour classifier based on a training set of size n , with resamples of size n' . In the infinite sampling case, under certain regularity conditions on the class distributions, the [excess risk](#) has the following asymptotic expansion^[5]

$$\mathcal{R}_{\mathcal{R}}(C_{n,n'}^{bnn}) - \mathcal{R}_{\mathcal{R}}(C^{Bayes}) = \left(B_1 \frac{n'}{n} + B_2 \frac{1}{(n')^{4/d}} \right) \{1 + o(1)\},$$

for some constants B_1 and B_2 . The optimal choice of n' , that balances the two terms in the asymptotic expansion, is given by $n' = Bn^{d/(d+4)}$ for some constant B .


History [\[edit\]](#)






Bagging (**B**ootstrap **a**ggregating) was proposed by [Leo Breiman](#) in 1994 to improve the classification by combining classifications of randomly generated training sets. See Breiman, 1994. Technical Report No. 421.

See also [\[edit\]](#)

- [Boosting \(meta-algorithm\)](#)
- [Bootstrapping \(statistics\)](#)
- [Cross-validation \(statistics\)](#)
- [Random forest](#)
- [Random subspace method](#) (attribute bagging)

References [\[edit\]](#)

- [↑] Aslam, Javed A.; Popa, Raluca A.; and Rivest, Ronald L. (2007); *On Estimating the Size and Confidence of a Statistical Audit* , Proceedings of the Electronic Voting Technology Workshop (EVT '07), Boston, MA, August 6, 2007. More generally, when drawing with replacement n' values out of a set of n (different and equally likely), the expected number of unique draws is $n(1 - e^{-n'/n})$.
- [↑] Sahu, A., Runger, G., Apley, D., Image denoising with a multi-phase kernel principal component approach and an ensemble version, IEEE Applied Imagery Pattern Recognition Workshop, pp.1-7, 2011.
- [↑] Shinde, Amit, Anshuman Sahu, Daniel Apley, and George Runger. "Preimages for Variation Patterns from Kernel PCA and Bagging." IIE Transactions, Vol.46, Iss.5, 2014

4. [^] Castelli, Vittorio. "Nearest Neighbor Classifiers, p.5"  (PDF). *columbia.edu*. Columbia University. Retrieved 25 April 2015.
5. [^] Samworth R. J. (2012). "Optimal weighted nearest neighbour classifiers". *Annals of Statistics* **40** (5): 2733–2763. doi:10.1214/12-AOS1049 .
- Breiman, Leo (1996). "Bagging predictors". *Machine Learning* **24** (2): 123–140. doi:10.1007/BF00058655 . CiteSeerX: 10.1.1.32.9399 .
- Alfaro, E., Gámez, M. and García, N. (2012). "adabag: An R package for classification with AdaBoost.M1, AdaBoost-SAMME and Bagging" .

Categories: [Ensemble learning](#) | [Machine learning algorithms](#) | [Computational statistics](#)

This page was last modified on 23 July 2015, at 17:28.

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.

[Privacy policy](#) [About Wikipedia](#) [Disclaimers](#) [Contact Wikipedia](#) [Developers](#) [Mobile view](#)

