Bootstrap aggregating

**Bootstrap aggregating**, also called **bagging**, is a [machine learning ensemble](https://en.wikipedia.org/wiki/Ensemble_learning) [meta-algorithm](https://en.wikipedia.org/wiki/Meta-algorithm) designed to improve the stability and accuracy of [machine learning](https://en.wikipedia.org/wiki/Machine_learning) algorithms used in [statistical classification](https://en.wikipedia.org/wiki/Statistical_classification) and [regression](https://en.wikipedia.org/wiki/Regression_analysis). It also reduces [variance](https://en.wikipedia.org/wiki/Variance) and helps to avoid [overfitting](https://en.wikipedia.org/wiki/Overfitting" \o "Overfitting). Although it is usually applied to [decision tree](https://en.wikipedia.org/wiki/Decision_tree_learning) methods, it can be used with any type of method. Bagging is a special case of the [model averaging](https://en.wikipedia.org/wiki/Ensemble_learning) approach.

History

Bagging (**B**ootstrap **agg**regat**ing**) was proposed by [Leo Breiman](https://en.wikipedia.org/wiki/Leo_Breiman) in 1994 to improve classification by combining classifications of randomly generated training sets. See Breiman, 1994. Technical Report No. 421.

Description of the technique[[edit](https://en.wikipedia.org/w/index.php?title=Bootstrap_aggregating&action=edit&section=2" \o "Edit section: Description of the technique)]

Given a standard [training set](https://en.wikipedia.org/wiki/Training_set)  of size *n*, bagging generates *m* new training sets {\displaystyle D\_{i}}, each of size *n′*, by [sampling](https://en.wikipedia.org/wiki/Sampling_(statistics)" \o "Sampling (statistics))from *D* [uniformly](https://en.wikipedia.org/wiki/Probability_distribution#With_finite_support) and [with replacement](https://en.wikipedia.org/wiki/Sampling_(statistics)#Replacement_of_selected_units). By sampling with replacement, some observations may be repeated in each {\displaystyle D\_{i}}. If *n*[*′*](https://en.wikipedia.org/wiki/Prime_(symbol))=*n*, then for large *n* the set {\displaystyle D\_{i}} is expected to have the fraction (1 - 1/[*e*](https://en.wikipedia.org/wiki/E_(mathematical_constant))) (≈63.2%) of the unique examples of *D*, the rest being duplicates.[[1]](https://en.wikipedia.org/wiki/Bootstrap_aggregating#cite_note-1) This kind of sample is known as a [bootstrap](https://en.wikipedia.org/wiki/Bootstrap_(statistics)) sample. The *m* models are fitted using the above *m* bootstrap samples and combined by averaging the output (for regression) or voting (for classification).

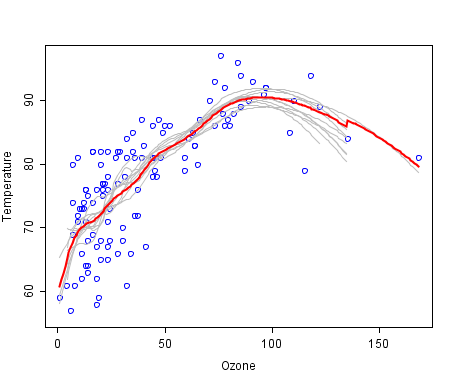
Bagging leads to "improvements for unstable procedures" (Breiman, 1996), which include, for example, [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_networks), [classification and regression trees](https://en.wikipedia.org/wiki/Classification_and_regression_tree), and subset selection in [linear regression](https://en.wikipedia.org/wiki/Linear_regression) (Breiman, 1994). An interesting application of bagging showing improvement in preimage learning is provided here.[[2]](https://en.wikipedia.org/wiki/Bootstrap_aggregating#cite_note-2)[[3]](https://en.wikipedia.org/wiki/Bootstrap_aggregating#cite_note-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](https://en.wikipedia.org/w/index.php?title=Bootstrap_aggregating&action=edit&section=3" \o "Edit section: Example: Ozone data)]

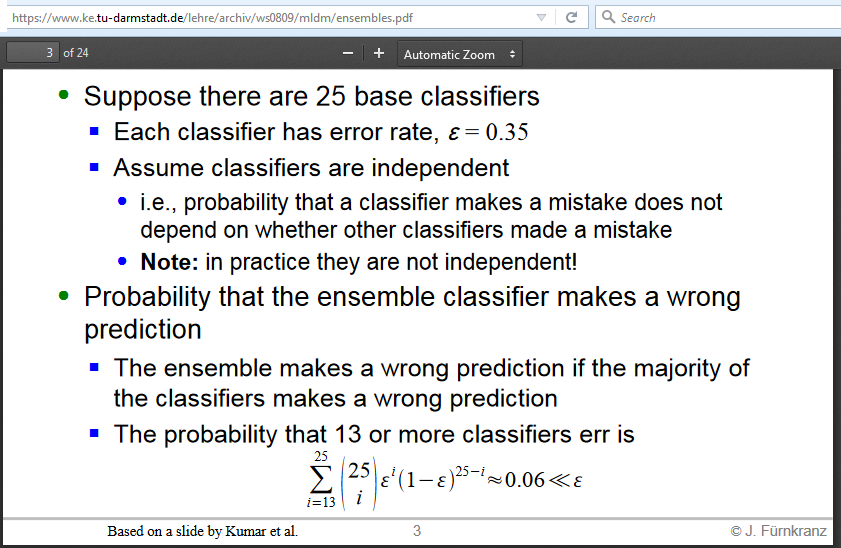
To illustrate the basic principles of bagging, below is an analysis on the relationship between [ozone](https://en.wikipedia.org/wiki/Ozone) and temperature (data from [Rousseeuw](https://en.wikipedia.org/wiki/Peter_Rousseeuw" \o "Peter Rousseeuw) and Leroy (1986), analysis done in [R](https://en.wikipedia.org/wiki/R_(programming_language))).

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](https://en.wikipedia.org/wiki/Local_regression) smoothers (with span 0.5) are used. Instead of building a single smoother from the complete data set, 100 [bootstrap](https://en.wikipedia.org/wiki/Bootstrap_(statistics)) 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 bagged predictor (red line). Clearly, the mean is more stable and there is less [overfit](https://en.wikipedia.org/wiki/Overfitting" \o "Overfitting).

[](https://en.wikipedia.org/wiki/File:Ozone.png)

……………………… ……………………… ……………………… ……………………… ……………………… ……………………



……………………… ……………………… ……………………… ……………………… ……………………… ……………………