

Random forest

Random forests or **random decision forests** is an [ensemble learning](#) method for [classification](#), [regression](#) and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned.^{[1][2]} Random decision forests correct for decision trees' habit of [overfitting](#) to their [training set](#).^{[3]:587–588} Random forests generally outperform [decision trees](#), but their accuracy is lower than [gradient boosted trees](#). However, data characteristics can affect their performance.^{[4][5]}

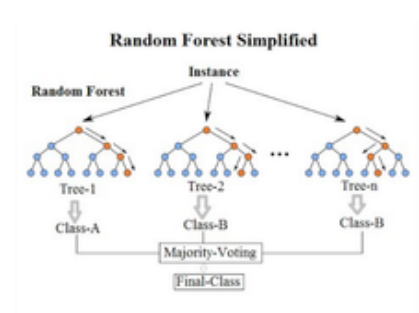


Diagram of a random decision forest

The first algorithm for random decision forests was created in 1995 by [Tin Kam Ho](#)^[1] using the [random subspace method](#),^[2] which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.^{[6][7][8]}

An extension of the algorithm was developed by [Leo Breiman](#)^[9] and [Adele Cutler](#),^[10] who registered^[11] "Random Forests" as a [trademark](#) in 2006 (as of 2019, owned by [Minitab, Inc.](#)).^[12] The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho^[1] and later independently by [Amit and Geman](#)^[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

Contents

History

Algorithm

- [Preliminaries: decision tree learning](#)
- [Bagging](#)
- [From bagging to random forests](#)
- [ExtraTrees](#)

Properties

- [Variable importance](#)
- [Relationship to nearest neighbors](#)

Unsupervised learning with random forests

Variants

Kernel random forest

- [History](#)
- [Notations and definitions](#)
 - [Preliminaries: Centered forests](#)
 - [Uniform forest](#)
 - [From random forest to KeRF](#)
 - [Centered KeRF](#)
 - [Uniform KeRF](#)

Properties

- [Relation between KeRF and random forest](#)

Relation between infinite KeRF and infinite random forest

Consistency results

Consistency of centered KeRF

Consistency of uniform KeRF

Disadvantages

See also

References

Further reading

External links

History

The general method of random decision forests was first proposed by Ho in 1995.^[1] Ho established that forests of trees splitting with oblique hyperplanes can gain accuracy as they grow without suffering from overtraining, as long as the forests are randomly restricted to be sensitive to only selected feature dimensions. A subsequent work along the same lines^[2] concluded that other splitting methods behave similarly, as long as they are randomly forced to be insensitive to some feature dimensions. Note that this observation of a more complex classifier (a larger forest) getting more accurate nearly monotonically is in sharp contrast to the common belief that the complexity of a classifier can only grow to a certain level of accuracy before being hurt by overfitting. The explanation of the forest method's resistance to overtraining can be found in Kleinberg's theory of stochastic discrimination.^{[6][7][8]}

The early development of Breiman's notion of random forests was influenced by the work of Amit and Geman^[13] who introduced the idea of searching over a random subset of the available decisions when splitting a node, in the context of growing a single tree. The idea of random subspace selection from Ho^[2] was also influential in the design of random forests. In this method a forest of trees is grown, and variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each tree or each node. Finally, the idea of randomized node optimization, where the decision at each node is selected by a randomized procedure, rather than a deterministic optimization was first introduced by Thomas G. Dietterich.^[14]

The proper introduction of random forests was made in a paper by Leo Breiman.^[9] This paper describes a method of building a forest of uncorrelated trees using a CART like procedure, combined with randomized node optimization and bagging. In addition, this paper combines several ingredients, some previously known and some novel, which form the basis of the modern practice of random forests, in particular:

1. Using out-of-bag error as an estimate of the generalization error.
2. Measuring variable importance through permutation.

The report also offers the first theoretical result for random forests in the form of a bound on the generalization error which depends on the strength of the trees in the forest and their correlation.

Algorithm

Preliminaries: decision tree learning

Decision trees are a popular method for various machine learning tasks. Tree learning "come[s] closest to meeting the requirements for serving as an off-the-shelf procedure for data mining", say Hastie et al., "because it is invariant under scaling and various other transformations of feature values, is robust to inclusion of irrelevant features, and produces inspectable models. However, they are seldom accurate".^{[3]:352}

In particular, trees that are grown very deep tend to learn highly irregular patterns: they overfit their training sets, i.e. have low bias, but very high variance. Random forests are a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance.^{[3]:587–588} This comes at the expense of a small increase in the bias and some loss of interpretability, but generally greatly boosts the performance in the final model.

Forests are like the pulling together of decision tree algorithm efforts. Taking the teamwork of many trees thus improving the performance of a single random tree. Though not quite similar, forests give the effects of a k-fold cross validation.

Bagging

The training algorithm for random forests applies the general technique of bootstrap aggregating, or bagging, to tree learners. Given a training set $X = x_1, \dots, x_n$ with responses $Y = y_1, \dots, y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples:

For $b = 1, \dots, B$:

1. Sample, with replacement, n training examples from X, Y ; call these X_b, Y_b .
2. Train a classification or regression tree f_b on X_b, Y_b .

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x' :

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x')$$

or by taking the majority vote in the case of classification trees.

This bootstrapping procedure leads to better model performance because it decreases the variance of the model, without increasing the bias. This means that while the predictions of a single tree are highly sensitive to noise in its training set, the average of many trees is not, as long as the trees are not correlated. Simply training many trees on a single training set would give strongly correlated trees (or even the same tree many times, if the training algorithm is deterministic); bootstrap sampling is a way of de-correlating the trees by showing them different training sets.

Additionally, an estimate of the uncertainty of the prediction can be made as the standard deviation of the predictions from all the individual regression trees on x' :

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - \hat{f})^2}{B - 1}}.$$

The number of samples/trees, B , is a free parameter. Typically, a few hundred to several thousand trees are used, depending on the size and nature of the training set. An optimal number of trees B can be found using cross-validation, or by observing the out-of-bag error: the mean prediction error on each training sample x_i , using only the trees that did not have x_i in their bootstrap sample.^[15] The training and test error tend to level off after some number of trees have been fit.

From bagging to random forests

The above procedure describes the original bagging algorithm for trees. Random forests also include another type of bagging scheme: they use a modified tree learning algorithm that selects, at each candidate split in the learning process, a random subset of the features. This process is sometimes called "feature bagging". The reason for doing this is the correlation of the trees in an ordinary bootstrap sample: if one or a few features are very strong predictors

for the response variable (target output), these features will be selected in many of the B trees, causing them to become correlated. An analysis of how bagging and random subspace projection contribute to accuracy gains under different conditions is given by Ho.^[16]

Typically, for a classification problem with p features, \sqrt{p} (rounded down) features are used in each split.^{[3]:592} For regression problems the inventors recommend $p/3$ (rounded down) with a minimum node size of 5 as the default.^{[3]:592} In practice the best values for these parameters will depend on the problem, and they should be treated as tuning parameters.^{[3]:592}

ExtraTrees

Adding one further step of randomization yields *extremely randomized trees*, or ExtraTrees. While similar to ordinary random forests in that they are an ensemble of individual trees, there are two main differences: first, each tree is trained using the whole learning sample (rather than a bootstrap sample), and second, the top-down splitting in the tree learner is randomized. Instead of computing the locally *optimal* cut-point for each feature under consideration (based on, e.g., information gain or the Gini impurity), a *random* cut-point is selected. This value is selected from a uniform distribution within the feature's empirical range (in the tree's training set). Then, of all the randomly generated splits, the split that yields the highest score is chosen to split the node. Similar to ordinary random forests, the number of randomly selected features to be considered at each node can be specified. Default values for this parameter are \sqrt{p} for classification and p for regression, where p is the number of features in the model.^[17]

Properties

Variable importance

Random forests can be used to rank the importance of variables in a regression or classification problem in a natural way. The following technique was described in Breiman's original paper^[9] and is implemented in the R package *randomForest*.^[10]

The first step in measuring the variable importance in a data set $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1}^n$ is to fit a random forest to the data. During the fitting process the out-of-bag error for each data point is recorded and averaged over the forest (errors on an independent test set can be substituted if bagging is not used during training).

To measure the importance of the j -th feature after training, the values of the j -th feature are permuted among the training data and the out-of-bag error is again computed on this perturbed data set. The importance score for the j -th feature is computed by averaging the difference in out-of-bag error before and after the permutation over all trees. The score is normalized by the standard deviation of these differences.

Features which produce large values for this score are ranked as more important than features which produce small values. The statistical definition of the variable importance measure was given and analyzed by Zhu *et al.*^[18]

This method of determining variable importance has some drawbacks. For data including categorical variables with different number of levels, random forests are biased in favor of those attributes with more levels. Methods such as partial permutations^{[19][20][4]} and growing unbiased trees^{[21][22]} can be used to solve the problem. If the data contain groups of correlated features of similar relevance for the output, then smaller groups are favored over larger groups.^[23]

Relationship to nearest neighbors

A relationship between random forests and the k -nearest neighbor algorithm (k -NN) was pointed out by Lin and Jeon in 2002.^[24] It turns out that both can be viewed as so-called *weighted neighborhoods schemes*. These are models built from a training set $\{(x_i, y_i)\}_{i=1}^n$ that make predictions \hat{y} for new points x' by looking at the "neighborhood" of the point, formalized by a weight function W :

$$\hat{y} = \sum_{i=1}^n W(\mathbf{x}_i, \mathbf{x}') y_i.$$

Here, $W(\mathbf{x}_i, \mathbf{x}')$ is the non-negative weight of the i 'th training point relative to the new point \mathbf{x}' in the same tree. For any particular \mathbf{x}' , the weights for points \mathbf{x}_i must sum to one. Weight functions are given as follows:

- In k -NN, the weights are $W(\mathbf{x}_i, \mathbf{x}') = \frac{1}{k}$ if \mathbf{x}_i is one of the k points closest to \mathbf{x}' , and zero otherwise.
- In a tree, $W(\mathbf{x}_i, \mathbf{x}') = \frac{1}{k'}$ if \mathbf{x}_i is one of the k' points in the same leaf as \mathbf{x}' , and zero otherwise.

Since a forest averages the predictions of a set of m trees with individual weight functions W_j , its predictions are

$$\hat{y} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n W_j(\mathbf{x}_i, \mathbf{x}') y_i = \sum_{i=1}^n \left(\frac{1}{m} \sum_{j=1}^m W_j(\mathbf{x}_i, \mathbf{x}') \right) y_i.$$

This shows that the whole forest is again a weighted neighborhood scheme, with weights that average those of the individual trees. The neighbors of \mathbf{x}' in this interpretation are the points \mathbf{x}_i sharing the same leaf in any tree j . In this way, the neighborhood of \mathbf{x}' depends in a complex way on the structure of the trees, and thus on the structure of the training set. Lin and Jeon show that the shape of the neighborhood used by a random forest adapts to the local importance of each feature.^[24]

Unsupervised learning with random forests

As part of their construction, random forest predictors naturally lead to a dissimilarity measure among the observations. One can also define a random forest dissimilarity measure between unlabeled data: the idea is to construct a random forest predictor that distinguishes the "observed" data from suitably generated synthetic data.^{[9][25]} The observed data are the original unlabeled data and the synthetic data are drawn from a reference distribution. A random forest dissimilarity can be attractive because it handles mixed variable types very well, is invariant to monotonic transformations of the input variables, and is robust to outlying observations. The random forest dissimilarity easily deals with a large number of semi-continuous variables due to its intrinsic variable selection; for example, the "Addcl 1" random forest dissimilarity weighs the contribution of each variable according to how dependent it is on other variables. The random forest dissimilarity has been used in a variety of applications, e.g. to find clusters of patients based on tissue marker data.^[26]

Variants

Instead of decision trees, linear models have been proposed and evaluated as base estimators in random forests, in particular multinomial logistic regression and naive Bayes classifiers.^{[5][27][28]} In cases that the relationship between the predictors and the target variable is linear, the base learners may have an equally high accuracy as the ensemble learner.^{[29][5]}

Kernel random forest

In machine learning, kernel random forests (KeRF) establish the connection between random forests and kernel methods. By slightly modifying their definition, random forests can be rewritten as kernel methods, which are more interpretable and easier to analyze.^[30]

History

Leo Breiman^[31] was the first person to notice the link between random forest and kernel methods. He pointed out that random forests which are grown using i.i.d. random vectors in the tree construction are equivalent to a kernel acting on the true margin. Lin and Jeon^[32] established the connection between random forests and adaptive nearest neighbor, implying that random forests can be seen as adaptive kernel estimates. Davies and Ghahramani^[33] proposed Random Forest Kernel and show that it can empirically outperform state-of-art kernel methods. Scornet^[30] first defined KeRF estimates and gave the explicit link between KeRF estimates and random forest. He also gave explicit expressions for kernels based on centered random forest^[34] and uniform random forest,^[35] two simplified models of random forest. He named these two KeRFs Centered KeRF and Uniform KeRF, and proved upper bounds on their rates of consistency.

Notations and definitions

Preliminaries: Centered forests

Centered forest^[34] is a simplified model for Breiman's original random forest, which uniformly selects an attribute among all attributes and performs splits at the center of the cell along the pre-chosen attribute. The algorithm stops when a fully binary tree of level k is built, where $k \in \mathbb{N}$ is a parameter of the algorithm.

Uniform forest

Uniform forest^[35] is another simplified model for Breiman's original random forest, which uniformly selects a feature among all features and performs splits at a point uniformly drawn on the side of the cell, along the preselected feature.

From random forest to KeRF

Given a training sample $\mathcal{D}_n = \{(\mathbf{X}_i, Y_i)\}_{i=1}^n$ of $[0, 1]^p \times \mathbb{R}$ -valued independent random variables distributed as the independent prototype pair (\mathbf{X}, Y) , where $\mathbb{E}[Y^2] < \infty$. We aim at predicting the response Y , associated with the random variable \mathbf{X} , by estimating the regression function $m(\mathbf{x}) = \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$. A random regression forest is an ensemble of M randomized regression trees. Denote $m_n(\mathbf{x}, \Theta_j)$ the predicted value at point \mathbf{x} by the j -th tree, where $\Theta_1, \dots, \Theta_M$ are independent random variables, distributed as a generic random variable Θ , independent of the sample \mathcal{D}_n . This random variable can be used to describe the randomness induced by node splitting and the sampling procedure for tree construction. The trees are combined to form the finite forest estimate

$$m_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{1}{M} \sum_{j=1}^M m_n(\mathbf{x}, \Theta_j). \text{ For regression trees, we have } m_n = \sum_{i=1}^n \frac{Y_i \mathbf{1}_{\mathbf{x}_i \in A_n(\mathbf{x}, \Theta_j)}}{N_n(\mathbf{x}, \Theta_j)}, \text{ where } A_n(\mathbf{x}, \Theta_j) \text{ is the cell containing } \mathbf{x}, \text{ designed with randomness } \Theta_j \text{ and dataset } \mathcal{D}_n, \text{ and } N_n(\mathbf{x}, \Theta_j) = \sum_{i=1}^n \mathbf{1}_{\mathbf{x}_i \in A_n(\mathbf{x}, \Theta_j)}.$$

Thus random forest estimates satisfy, for all $\mathbf{x} \in [0, 1]^d$,

$$m_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{1}{M} \sum_{j=1}^M \left(\sum_{i=1}^n \frac{Y_i \mathbf{1}_{\mathbf{x}_i \in A_n(\mathbf{x}, \Theta_j)}}{N_n(\mathbf{x}, \Theta_j)} \right). \text{ Random regression forest has two levels of}$$

averaging, first over the samples in the target cell of a tree, then over all trees. Thus the contributions of observations that are in cells with a high density of data points are smaller than that of observations which belong to less populated cells. In order to improve the random forest methods and compensate the misestimation, Scornet^[30] defined KeRF by

$$\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{1}{\sum_{j=1}^M N_n(\mathbf{x}, \Theta_j)} \sum_{j=1}^M \sum_{i=1}^n Y_i \mathbf{1}_{\mathbf{x}_i \in A_n(\mathbf{x}, \Theta_j)},$$

which is equal to the mean of the Y_i 's falling in the cells containing \mathbf{x} in the forest. If we define the connection function of the M finite forest as $K_{M,n}(\mathbf{x}, \mathbf{z}) = \frac{1}{M} \sum_{j=1}^M \mathbf{1}_{\mathbf{z} \in A_n(\mathbf{x}, \Theta_j)}$, i.e. the proportion of cells shared between \mathbf{x} and \mathbf{z} , then almost surely we have $\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M) = \frac{\sum_{i=1}^n Y_i K_{M,n}(\mathbf{x}, \mathbf{x}_i)}{\sum_{\ell=1}^n K_{M,n}(\mathbf{x}, \mathbf{x}_\ell)}$, which defines the KeRF.

Centered KeRF

The construction of Centered KeRF of level k is the same as for centered forest, except that predictions are made by $\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M)$, the corresponding kernel function, or connection function is

$$K_k^{cc}(\mathbf{x}, \mathbf{z}) = \sum_{k_1, \dots, k_d, \sum_{j=1}^d k_j = k} \frac{k!}{k_1! \dots k_d!} \left(\frac{1}{d}\right)^k \prod_{j=1}^d \mathbf{1}_{\lceil 2^{k_j} x_j \rceil = \lceil 2^{k_j} z_j \rceil},$$

for all $\mathbf{x}, \mathbf{z} \in [0, 1]^d$.

Uniform KeRF

Uniform KeRF is built in the same way as uniform forest, except that predictions are made by $\tilde{m}_{M,n}(\mathbf{x}, \Theta_1, \dots, \Theta_M)$, the corresponding kernel function, or connection function is

$$K_k^{uf}(\mathbf{0}, \mathbf{x}) = \sum_{k_1, \dots, k_d, \sum_{j=1}^d k_j = k} \frac{k!}{k_1! \dots k_d!} \left(\frac{1}{d}\right)^k \prod_{m=1}^d \left(1 - |x_m| \sum_{j=0}^{k_m-1} \frac{(-\ln |x_m|)^j}{j!}\right) \text{ for all } \mathbf{x} \in [0, 1]^d.$$

Properties

Relation between KeRF and random forest

Predictions given by KeRF and random forests are close if the number of points in each cell is controlled:

Assume that there exist sequences $(a_n), (b_n)$ such that, almost surely,

$$a_n \leq N_n(\mathbf{x}, \Theta) \leq b_n \text{ and } a_n \leq \frac{1}{M} \sum_{m=1}^M N_n \mathbf{x}, \Theta_m \leq b_n.$$

Then almost surely,

$$|m_{M,n}(\mathbf{x}) - \tilde{m}_{M,n}(\mathbf{x})| \leq \frac{b_n - a_n}{a_n} \tilde{m}_{M,n}(\mathbf{x}).$$

Relation between infinite KeRF and infinite random forest

When the number of trees M goes to infinity, then we have infinite random forest and infinite KeRF. Their estimates are close if the number of observations in each cell is bounded:

Assume that there exist sequences $(\varepsilon_n), (a_n), (b_n)$ such that, almost surely

- $\mathbb{E}[N_n(\mathbf{x}, \Theta)] \geq 1$,
- $\mathbb{P}[a_n \leq N_n(\mathbf{x}, \Theta) \leq b_n \mid \mathcal{D}_n] \geq 1 - \varepsilon_n/2$,
- $\mathbb{P}[a_n \leq \mathbb{E}_\Theta[N_n(\mathbf{x}, \Theta)] \leq b_n \mid \mathcal{D}_n] \geq 1 - \varepsilon_n/2$,

Then almost surely,

$$|m_{\infty,n}(\mathbf{x}) - \tilde{m}_{\infty,n}(\mathbf{x})| \leq \frac{b_n - a_n}{a_n} \tilde{m}_{\infty,n}(\mathbf{x}) + n\varepsilon_n \left(\max_{1 \leq i \leq n} Y_i \right).$$

Consistency results

Assume that $\mathbf{Y} = \mathbf{m}(\mathbf{X}) + \varepsilon$, where ε is a centered Gaussian noise, independent of \mathbf{X} , with finite variance $\sigma^2 < \infty$. Moreover, \mathbf{X} is uniformly distributed on $[0, 1]^d$ and \mathbf{m} is Lipschitz. Scornet^[30] proved upper bounds on the rates of consistency for centered KeRF and uniform KeRF.

Consistency of centered KeRF

Providing $k \rightarrow \infty$ and $n/2^k \rightarrow \infty$, there exists a constant $C_1 > 0$ such that, for all n , $\mathbb{E}[\tilde{m}_n^{cc}(\mathbf{X}) - m(\mathbf{X})]^2 \leq C_1 n^{-1/(3+d \log 2)} (\log n)^2$.

Consistency of uniform KeRF

Providing $k \rightarrow \infty$ and $n/2^k \rightarrow \infty$, there exists a constant $C > 0$ such that, $\mathbb{E}[\tilde{m}_n^{uf}(\mathbf{X}) - m(\mathbf{X})]^2 \leq C n^{-2/(6+3d \log 2)} (\log n)^2$.

Disadvantages

While random forests often achieve higher accuracy than a single decision tree, they sacrifice the intrinsic interpretability present in decision trees. Decision trees are among a fairly small family of machine learning models that are easily interpretable along with linear models, rule-based models, and attention-based models. This interpretability is one of the most desirable qualities of decision trees. It allows developers to confirm that the model has learned realistic information from the data and allows end-users to have trust and confidence in the decisions made by the model.^{[5][31]} For example, following the path that a decision tree takes to make its decision is quite trivial, but following the paths of tens or hundreds of trees is much harder. To achieve both performance and interpretability, some model compression techniques allow transforming a random forest into a minimal "born-again" decision tree that faithfully reproduces the same decision function.^{[5][36][37]} If it is established that the predictive attributes are linearly correlated with the target variable, using random forest may not enhance the accuracy of the base learner.^{[5][29]} Furthermore, in problems with multiple categorical variables, random forest may not be able to increase the accuracy of the base learner.^[38]

See also

- Boosting – Method in machine learning
- Decision tree learning – Machine learning algorithm
- Ensemble learning – Statistics and machine learning technique
- Gradient boosting – Machine learning technique
- Non-parametric statistics
- Randomized algorithm – Algorithm that employs a degree of randomness as part of its logic or procedure

References

1. Ho, Tin Kam (1995). *Random Decision Forests* (<https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>) (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (<http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf>) (PDF) on 17 April 2016. Retrieved 5 June 2016.
2. Ho TK (1998). "The Random Subspace Method for Constructing Decision Forests" (<http://ect.bell-labs.com/who/tkh/publications/papers/df.pdf>) (PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **20** (8): 832–844. doi:10.1109/34.709601 (<https://doi.org/10.1109%2F34.709601>).
3. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2008). *The Elements of Statistical Learning* (<http://www-stat.stanford.edu/~tibs/ElemStatLearn/>) (2nd ed.). Springer. ISBN 0-387-95284-5.
4. Pirayonesi S. Madeh; El-Diraby Tamer E. (2020-06-01). "Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems". *Journal of Transportation Engineering, Part B: Pavements*. **146** (2): 04020022. doi:10.1061/JPEODX.0000175 (<https://doi.org/10.1061%2FJPEODX.0000175>). S2CID 216485629 (<https://api.semanticscholar.org/CorpusID:216485629>).
5. Pirayonesi, S. Madeh; El-Diraby, Tamer E. (2021-02-01). "Using Machine Learning to Examine Impact of Type of Performance Indicator on Flexible Pavement Deterioration Modeling" (<http://ascelibrary.org/doi/10.1061/%28ASCE%29IS.1943-555X.0000602>). *Journal of Infrastructure Systems*. **27** (2): 04021005. doi:10.1061/(ASCE)IS.1943-555X.0000602 (<https://doi.org/10.1061%2F%28ASCE%29IS.1943-555X.0000602>). ISSN 1076-0342 (<https://www.worldcat.org/issn/1076-0342>). S2CID 233550030 (<https://api.semanticscholar.org/CorpusID:233550030>).
6. Kleinberg E (1990). "Stochastic Discrimination" (<https://web.archive.org/web/20180118124007/http://pdfs.semanticscholar.org/faa4/c502a824a9d64bf3dc26eb90a2c32367921f.pdf>) (PDF). *Annals of Mathematics and Artificial Intelligence*. **1** (1–4): 207–239. CiteSeerX 10.1.1.25.6750 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.25.6750>). doi:10.1007/BF01531079 (<https://doi.org/10.1007%2FBF01531079>). S2CID 206795835 (<https://api.semanticscholar.org/CorpusID:206795835>). Archived from the original (<https://pdfs.semanticscholar.org/faa4/c502a824a9d64bf3dc26eb90a2c32367921f.pdf>) (PDF) on 2018-01-18.
7. Kleinberg E (1996). "An Overtraining-Resistant Stochastic Modeling Method for Pattern Recognition" (<https://doi.org/10.1214%2Faos%2F1032181157>). *Annals of Statistics*. **24** (6): 2319–2349. doi:10.1214/aos/1032181157 (<https://doi.org/10.1214%2Faos%2F1032181157>). MR 1425956 (<http://www.ams.org/mathscinet-getitem?mr=1425956>).
8. Kleinberg E (2000). "On the Algorithmic Implementation of Stochastic Discrimination" (<https://web.archive.org/web/20180118124006/https://pdfs.semanticscholar.org/8956/845b0701ec57094c7a8b4ab1f41386899aea.pdf>) (PDF). *IEEE Transactions on PAMI*. **22** (5): 473–490. CiteSeerX 10.1.1.33.4131 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.33.4131>). doi:10.1109/34.857004 (<https://doi.org/10.1109%2F34.857004>). S2CID 3563126 (<https://api.semanticscholar.org/CorpusID:3563126>). Archived from the original (<https://pdfs.semanticscholar.org/8956/845b0701ec57094c7a8b4ab1f41386899aea.pdf>) (PDF) on 2018-01-18.
9. Breiman L (2001). "Random Forests" (<https://doi.org/10.1023%2FA%3A1010933404324>). *Machine Learning*. **45** (1): 5–32. Bibcode:2001MachL..45....5B (<https://ui.adsabs.harvard.edu/abs/2001MachL..45....5B>). doi:10.1023/A:1010933404324 (<https://doi.org/10.1023%2FA%3A1010933404324>).
10. Liaw A (16 October 2012). "Documentation for R package randomForest" (<https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>) (PDF). Retrieved 15 March 2013.
11. U.S. trademark registration number 3185828, registered 2006/12/19.
12. "RANDOM FORESTS Trademark of Health Care Productivity, Inc. - Registration Number 3185828 - Serial Number 78642027 :: Justia Trademarks" (<https://trademarks.justia.com/786/42/random-78642027.html>).
13. Amit Y, Geman D (1997). "Shape quantization and recognition with randomized trees" (http://www.cis.jhu.edu/publications/papers_in_database/GEMAN/shape.pdf) (PDF). *Neural Computation*. **9** (7): 1545–1588. CiteSeerX 10.1.1.57.6069 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.57.6069>). doi:10.1162/neco.1997.9.7.1545 (<https://doi.org/10.1162%2Fneco.1997.9.7.1545>). S2CID 12470146 (<https://api.semanticscholar.org/CorpusID:12470146>).

14. Dietterich, Thomas (2000). "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization" (<https://doi.org/10.1023/A:1007607513941>). *Machine Learning*. **40** (2): 139–157. doi:10.1023/A:1007607513941 (<https://doi.org/10.1023/A:1007607513941>).
15. Gareth James; Daniela Witten; Trevor Hastie; Robert Tibshirani (2013). *An Introduction to Statistical Learning* (<http://www-bcf.usc.edu/~gareth/ISL/>). Springer. pp. 316–321.
16. Ho, Tin Kam (2002). "A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors" (<http://ect.bell-labs.com/who/tkh/publications/papers/compare.pdf>) (PDF). *Pattern Analysis and Applications*. **5** (2): 102–112. doi:10.1007/s100440200009 (<https://doi.org/10.1007/s100440200009>). S2CID 7415435 (<https://api.semanticscholar.org/CorpusID:7415435>).
17. Geurts P, Ernst D, Wehenkel L (2006). "Extremely randomized trees" (<http://orbi.ulg.ac.be/bitstream/2268/9357/1/geurts-mlj-advance.pdf>) (PDF). *Machine Learning*. **63**: 3–42. doi:10.1007/s10994-006-6226-1 (<https://doi.org/10.1007/s10994-006-6226-1>).
18. Zhu R, Zeng D, Kosorok MR (2015). "Reinforcement Learning Trees" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4760114>). *Journal of the American Statistical Association*. **110** (512): 1770–1784. doi:10.1080/01621459.2015.1036994 (<https://doi.org/10.1080/01621459.2015.1036994>). PMC 4760114 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4760114>). PMID 26903687 (<https://pubmed.ncbi.nlm.nih.gov/26903687>).
19. Deng, H.; Runger, G.; Tuv, E. (2011). *Bias of importance measures for multi-valued attributes and solutions* (<https://www.researchgate.net/publication/221079908>). Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN). pp. 293–300.
20. Altmann A, Tološi L, Sander O, Lengauer T (May 2010). "Permutation importance: a corrected feature importance measure" (<https://doi.org/10.1093/bioinformatics/btq134>). *Bioinformatics*. **26** (10): 1340–7. doi:10.1093/bioinformatics/btq134 (<https://doi.org/10.1093/bioinformatics/btq134>). PMID 20385727 (<https://pubmed.ncbi.nlm.nih.gov/20385727>).
21. Strobl C, Boulesteix A, Augustin T (2007). "Unbiased split selection for classification trees based on the Gini index" (https://epub.uni-muenchen.de/1833/1/paper_464.pdf) (PDF). *Computational Statistics & Data Analysis*. **52**: 483–501. CiteSeerX 10.1.1.525.3178 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.525.3178>). doi:10.1016/j.csda.2006.12.030 (<https://doi.org/10.1016/j.csda.2006.12.030>).
22. Painsky A, Rosset S (2017). "Cross-Validated Variable Selection in Tree-Based Methods Improves Predictive Performance". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **39** (11): 2142–2153. arXiv:1512.03444 (<https://arxiv.org/abs/1512.03444>). doi:10.1109/tpami.2016.2636831 (<https://doi.org/10.1109/tpami.2016.2636831>). PMID 28114007 (<https://pubmed.ncbi.nlm.nih.gov/28114007>). S2CID 5381516 (<https://api.semanticscholar.org/CorpusID:5381516>).
23. Tolosi L, Lengauer T (July 2011). "Classification with correlated features: unreliability of feature ranking and solutions" (<https://doi.org/10.1093/bioinformatics/btr300>). *Bioinformatics*. **27** (14): 1986–94. doi:10.1093/bioinformatics/btr300 (<https://doi.org/10.1093/bioinformatics/btr300>). PMID 21576180 (<https://pubmed.ncbi.nlm.nih.gov/21576180>).
24. Lin, Yi; Jeon, Yongho (2002). *Random forests and adaptive nearest neighbors* (Technical report). Technical Report No. 1055. University of Wisconsin. CiteSeerX 10.1.1.153.9168 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.153.9168>).
25. Shi, T., Horvath, S. (2006). "Unsupervised Learning with Random Forest Predictors". *Journal of Computational and Graphical Statistics*. **15** (1): 118–138. CiteSeerX 10.1.1.698.2365 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.698.2365>). doi:10.1198/106186006X94072 (<https://doi.org/10.1198/106186006X94072>). JSTOR 27594168 (<https://www.jstor.org/stable/27594168>). S2CID 245216 (<https://api.semanticscholar.org/CorpusID:245216>).
26. Shi T, Seligson D, Belldegrun AS, Palotie A, Horvath S (April 2005). "Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma" (<https://doi.org/10.1038/modpathol.3800322>). *Modern Pathology*. **18** (4): 547–57. doi:10.1038/modpathol.3800322 (<https://doi.org/10.1038/modpathol.3800322>). PMID 15529185 (<https://pubmed.ncbi.nlm.nih.gov/15529185>).
27. Prinzie, A., Van den Poel, D. (2008). "Random Forests for multiclass classification: Random MultiNomial Logit". *Expert Systems with Applications*. **34** (3): 1721–1732. doi:10.1016/j.eswa.2007.01.029 (<https://doi.org/10.1016/j.eswa.2007.01.029>).

28. Prinzie, Anita (2007). "Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB". In Roland Wagner; Norman Revell; Günther Pernul (eds.). *Database and Expert Systems Applications: 18th International Conference, DEXA 2007, Regensburg, Germany, September 3-7, 2007, Proceedings*. Lecture Notes in Computer Science. Vol. 4653. pp. 349–358. doi:10.1007/978-3-540-74469-6_35 (https://doi.org/10.1007%2F978-3-540-74469-6_35). ISBN 978-3-540-74467-2.
29. Smith, Paul F.; Ganesh, Siva; Liu, Ping (2013-10-01). "A comparison of random forest regression and multiple linear regression for prediction in neuroscience" (<https://linkinghub.elsevier.com/retrieve/pii/S0165027013003026>). *Journal of Neuroscience Methods*. **220** (1): 85–91. doi:10.1016/j.jneumeth.2013.08.024 (<https://doi.org/10.1016%2Fj.jneumeth.2013.08.024>). PMID 24012917 (<https://pubmed.ncbi.nlm.nih.gov/24012917>). S2CID 13195700 (<https://api.semanticscholar.org/CorpusID:13195700>).
30. Scornet, Erwan (2015). "Random forests and kernel methods". arXiv:1502.03836 (<https://arxiv.org/abs/1502.03836>) [math.ST (<https://arxiv.org/archive/math>.ST)].
31. Breiman, Leo (2000). "Some infinity theory for predictor ensembles" (<https://statistics.berkeley.edu/tech-reports/579>). Technical Report 579, Statistics Dept. UCB.
32. Lin, Yi; Jeon, Yongho (2006). "Random forests and adaptive nearest neighbors". *Journal of the American Statistical Association*. **101** (474): 578–590. CiteSeerX 10.1.1.153.9168 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.153.9168>). doi:10.1198/016214505000001230 (<https://doi.org/10.1198%2F016214505000001230>). S2CID 2469856 (<https://api.semanticscholar.org/CorpusID:2469856>).
33. Davies, Alex; Ghahramani, Zoubin (2014). "The Random Forest Kernel and other kernels for big data from random partitions". arXiv:1402.4293 (<https://arxiv.org/abs/1402.4293>) [stat.ML (<https://arxiv.org/archive/stat>.ML)].
34. Breiman L, Ghahramani Z (2004). "Consistency for a simple model of random forests". *Statistical Department, University of California at Berkeley. Technical Report* (670). CiteSeerX 10.1.1.618.90 (<https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.618.90>).
35. Arlot S, Genuer R (2014). "Analysis of purely random forests bias". arXiv:1407.3939 (<https://arxiv.org/abs/1407.3939>) [math.ST (<https://arxiv.org/archive/math>.ST)].
36. Sagi, Omer; Rokach, Lior (2020). "Explainable decision forest: Transforming a decision forest into an interpretable tree" (<https://www.sciencedirect.com/science/article/pii/S1566253519307869>). *Information Fusion*. **61**: 124–138. doi:10.1016/j.inffus.2020.03.013 (<https://doi.org/10.1016%2Fj.inffus.2020.03.013>). S2CID 216444882 (<https://api.semanticscholar.org/CorpusID:216444882>).
37. Vidal, Thibaut; Schiffer, Maximilian (2020). "Born-Again Tree Ensembles" (<http://proceedings.mlr.press/v119/vidal20a.html>). *International Conference on Machine Learning*. PMLR. **119**: 9743–9753. arXiv:2003.11132 (<https://arxiv.org/abs/2003.11132>).
38. Pirayonesi, Sayed Madeh (November 2019). *Pirayonesi, S. M. (2019). The Application of Data Analytics to Asset Management: Deterioration and Climate Change Adaptation in Ontario Roads (Doctoral dissertation)* (<https://tspace.library.utoronto.ca/handle/1807/97601>) (Thesis).

Further reading

- Prinzie A, Poel D (2007). "Random Multiclass Classification: Generalizing Random Forests to Random MNL and Random NB" (<https://www.researchgate.net/publication/225175169>). *Database and Expert Systems Applications*. Lecture Notes in Computer Science. Vol. 4653. p. 349. doi:10.1007/978-3-540-74469-6_35 (https://doi.org/10.1007%2F978-3-540-74469-6_35). ISBN 978-3-540-74467-2.
- Denisko D, Hoffman MM (February 2018). "Classification and interaction in random forests" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5828645>). *Proceedings of the National Academy of Sciences of the United States of America*. **115** (8): 1690–1692. doi:10.1073/pnas.1800256115 (<https://doi.org/10.1073%2Fpnas.1800256115>). PMC 5828645 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5828645>). PMID 29440440 (<https://pubmed.ncbi.nlm.nih.gov/29440440>).

External links

- [Random Forests classifier description \(https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm\)](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm) (Leo Breiman's site)
 - [Liaw, Andy & Wiener, Matthew "Classification and Regression by randomForest" R News \(2002\) Vol. 2/3 p. 18 \(https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf\)](https://cran.r-project.org/doc/Rnews/Rnews_2002-3.pdf) (Discussion of the use of the random forest package for R)
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1094130824"

This page was last edited on 20 June 2022, at 23:00 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License 3.0; 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.