**Decision Trees**

<https://link.springer.com/chapter/10.1007/0-387-25465-X_9#citeas>

@Inbook{Rokach2005,

author="Rokach, Lior

and Maimon, Oded",

editor="Maimon, Oded

and Rokach, Lior",

title="Decision Trees",

bookTitle="Data Mining and Knowledge Discovery Handbook",

year="2005",

publisher="Springer US",

address="Boston, MA",

pages="165--192",

abstract="Decision Trees are considered to be one of the most popular approaches for representing classifiers. Researchers from various disciplines such as statistics, machine learning, pattern recognition, and Data Mining have dealt with the issue of growing a decision tree from available data. This paper presents an updated survey of current methods for constructing decision tree classifiers in a top-down manner. The chapter suggests a unified algorithmic framework for presenting these algorithms and describes various splitting criteria and pruning methodologies.",

isbn="978-0-387-25465-4",

doi="10.1007/0-387-25465-X\_9",

url="https://doi.org/10.1007/0-387-25465-X\_9"

}

**Random Forests**

[**https://link.springer.com/article/10.1023/A:1010933404324**](https://link.springer.com/article/10.1023/A:1010933404324)

TY - JOUR

AU - Breiman, Leo

PY - 2001

DA - 2001/10/01

TI - Random Forests

JO - Machine Learning

SP - 5

EP - 32

VL - 45

IS - 1

AB - Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favorably to Adaboost (Y. Freund & R. Schapire, Machine Learning: Proceedings of the Thirteenth International conference, \*\*\*, 148–156), but are more robust with respect to noise. Internal estimates monitor error, strength, and correlation and these are used to show the response to increasing the number of features used in the splitting. Internal estimates are also used to measure variable importance. These ideas are also applicable to regression.

SN - 1573-0565

UR - https://doi.org/10.1023/A:1010933404324

DO - 10.1023/A:1010933404324

ID - Breiman2001

ER -

**Extra Trees Classifier**

[**https://link.springer.com/article/10.1007/s10994-006-6226-1**](https://link.springer.com/article/10.1007/s10994-006-6226-1)

TY - JOUR

AU - Geurts, Pierre

AU - Ernst, Damien

AU - Wehenkel, Louis

PY - 2006

DA - 2006/04/01

TI - Extremely randomized trees

JO - Machine Learning

SP - 3

EP - 42

VL - 63

IS - 1

AB - This paper proposes a new tree-based ensemble method for supervised classification and regression problems. It essentially consists of randomizing strongly both attribute and cut-point choice while splitting a tree node. In the extreme case, it builds totally randomized trees whose structures are independent of the output values of the learning sample. The strength of the randomization can be tuned to problem specifics by the appropriate choice of a parameter. We evaluate the robustness of the default choice of this parameter, and we also provide insight on how to adjust it in particular situations. Besides accuracy, the main strength of the resulting algorithm is computational efficiency. A bias/variance analysis of the Extra-Trees algorithm is also provided as well as a geometrical and a kernel characterization of the models induced.

SN - 1573-0565

UR - https://doi.org/10.1007/s10994-006-6226-1

DO - 10.1007/s10994-006-6226-1

ID - Geurts2006

ER -

**k nearest neighbor**

<https://link.springer.com/chapter/10.1007/978-3-540-39964-3_62>

@InProceedings{10.1007/978-3-540-39964-3\_62,

author="Guo, Gongde

and Wang, Hui

and Bell, David

and Bi, Yaxin

and Greer, Kieran",

editor="Meersman, Robert

and Tari, Zahir

and Schmidt, Douglas C.",

title="KNN Model-Based Approach in Classification",

booktitle="On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE",

year="2003",

publisher="Springer Berlin Heidelberg",

address="Berlin, Heidelberg",

pages="986--996",

abstract="The k-Nearest-Neighbours (kNN) is a simple but effective method for classification. The major drawbacks with respect to kNN are (1) its low efficiency -- being a lazy learning method prohibits it in many applications such as dynamic web mining for a large repository, and (2) its dependency on the selection of a ``good value'' for k. In this paper, we propose a novel kNN type method for classification that is aimed at overcoming these shortcomings. Our method constructs a kNN model for the data, which replaces the data to serve as the basis of classification. The value of k is automatically determined, is varied for different data, and is optimal in terms of classification accuracy. The construction of the model reduces the dependency on k and makes classification faster. Experiments were carried out on some public datasets collected from the UCI machine learning repository in order to test our method. The experimental results show that the kNN based model compares well with C5.0 and kNN in terms of classification accuracy, but is more efficient than the standard kNN.",

isbn="978-3-540-39964-3"

}

**Logistic Regression**

<https://web.stanford.edu/~jurafsky/slp3/5.pdf>

**Support Vector Machine Classification**

<https://link.springer.com/chapter/10.1007/978-3-642-34041-3_27>

@ARTICLE{708428,

author={Hearst, M.A. and Dumais, S.T. and Osuna, E. and Platt, J. and Scholkopf, B.},

journal={IEEE Intelligent Systems and their Applications},

title={Support vector machines},

year={1998},

volume={13},

number={4},

pages={18-28},

doi={10.1109/5254.708428}}

**XGBoost**

<https://arxiv.org/abs/1603.02754>

@article{DBLP:journals/corr/ChenG16,

author = {Tianqi Chen and

Carlos Guestrin},

title = {XGBoost: {A} Scalable Tree Boosting System},

journal = {CoRR},

volume = {abs/1603.02754},

year = {2016},

url = {http://arxiv.org/abs/1603.02754},

eprinttype = {arXiv},

eprint = {1603.02754},

timestamp = {Mon, 13 Aug 2018 16:47:00 +0200},

biburl = {https://dblp.org/rec/journals/corr/ChenG16.bib},

bibsource = {dblp computer science bibliography, https://dblp.org}