Machine learning

From Wikipedia, the free encyclopedia

Machine learning is the subfield of computer science that, according to Arthur Samuel in 1959, gives "computers the ability to learn without being explicitly programmed."[1] Evolved from the study of pattern recognition and computational learning theory in artificial intelligence,[2] machine learning explores the study and construction of algorithms that can learn from and make predictions on data[3] – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions,[4]:2 through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach,[5] optical character recognition (OCR),[6] learning to rank, and computer vision.

Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining,[7] where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.[4]:vii[8] Machine learning can also be unsupervised[9] and be used to learn and establish baseline behavioral profiles for various entities[10] and then used to find meaningful anomalies.

Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data.[11]

As of 2016, machine learning is a buzzword, and according to the Gartner hype cycle of 2016, at its peak of inflated expectations.[12] Because finding patterns is hard, often not enough training data is available, and also because of the high expectations it often fails to deliver.[13][14]

Contents

Overview

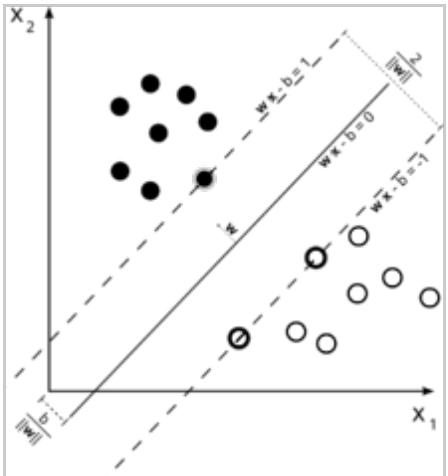
Tom M. Mitchell provided a widely quoted, more formal definition: "A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P* if its performance at tasks in *T*, as measured by *P*, improves with experience *E*."[15] This definition is notable for its defining machine learning in fundamentally operational rather than cognitive terms, thus following Alan Turing's proposal in his paper "Computing Machinery and Intelligence", that the question "Can machines think?" be replaced with the question "Can machines do what we (as thinking entities) can do?".[16] In the proposal he explores the various characteristics that could be possessed by a *thinking machine* and the various implications in constructing one.

Types of problems and tasks

Machine learning tasks are typically classified into three broad categories, depending on the nature of the learning "signal" or "feedback" available to a learning system. These are[17]

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent[4]:3). The program is provided feedback in terms of rewards and punishments as it navigates its problem space.

Between supervised and unsupervised learning is semi-supervised learning, where the teacher gives an incomplete training signal: a training set with some (often many) of the target outputs missing. Transduction is a special case of this principle where the entire set of problem instances is known at learning time, except that part of the targets are missing.


A support vector machine is a classifier that divides its input space into two regions, separated by a linear boundary. Here, it has learned to distinguish black and white circles.

Among other categories of machine learning problems, learning to learn learns its own inductive bias based on previous experience. Developmental learning, elaborated for robot learning, generates its own sequences (also called curriculum) of learning situations to cumulatively acquire repertoires of novel skills through autonomous self-exploration and social interaction with human teachers and using guidance mechanisms such as active learning, maturation, motor synergies, and imitation.

Another categorization of machine learning tasks arises when one considers the desired *output* of a machine-learned system:[4]:3

- In classification, inputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. This is typically tackled in a supervised way. Spam filtering is an example of classification, where the inputs are email (or other) messages and the classes are "spam" and "not spam".
- In regression, also a supervised problem, the outputs are continuous rather than discrete.
- In clustering, a set of inputs is to be divided into groups. Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task.
- Density estimation finds the distribution of inputs in some space.
- Dimensionality reduction simplifies inputs by mapping them into a lower-dimensional space. Topic modeling is a related problem, where a program is given a list of human language documents and is tasked to find out which documents cover similar topics.

History and relationships to other fields

As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics.[18] Probabilistic reasoning was also employed, especially in automated medical diagnosis.[17]:488

However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation.[17]:488 By 1980, expert systems had come to dominate AI, and statistics was out of favor.[19] Work on symbolic/knowledge-based learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval.[17]:708–710; 755 Neural networks research had been abandoned by AI and computer science around the same time. This line, too, was continued outside the AI/CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart and Hinton. Their main success came in the mid-1980s with the reinvention of backpropagation.[17]:25

Machine learning, reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, and toward methods and models borrowed from statistics and probability theory.[19] It also benefited from the increasing availability of digitized information, and the possibility to distribute that via the Internet.

Machine learning and data mining often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on *known* properties learned from the training data, data mining focuses on the discovery of (previously) *unknown* properties in the data (this is the analysis step of Knowledge Discovery in Databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy. Much of the confusion between these two research communities (which do often have separate conferences and separate journals, ECML PKDD being a major exception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability to *reproduce known* knowledge, while in Knowledge Discovery and Data Mining (KDD) the key task is the discovery of previously *unknown* knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the unavailability of training data.

Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set of examples). The difference between the two fields arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples.[20]

Relation to statistics

Machine learning and statistics are closely related fields. According to Michael I. Jordan, the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics.[21] He also suggested the term data science as a placeholder to call the overall field.[21]

Leo Breiman distinguished two statistical modelling paradigms: data model and algorithmic model,[22] wherein 'algorithmic model' means more or less the machine learning algorithms like Random forest.

Some statisticians have adopted methods from machine learning, leading to a combined field that they call *statistical learning*.[23]

Theory

A core objective of a learner is to generalize from its experience.[24][25] Generalization in this context is the ability of a learning machine to perform accurately on new, unseen examples/tasks after having experienced a learning data set. The training examples come from some generally unknown probability distribution (considered representative of the space of occurrences) and the learner has to build a general model about this space that enables it to produce sufficiently accurate predictions in new cases.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory. Because training sets are finite and the future is uncertain, learning theory usually does not yield guarantees of the performance of algorithms. Instead, probabilistic bounds on the performance are quite common. The bias–variance decomposition is one way to quantify generalization error.

For the best performance in the context of generalization, the complexity of the hypothesis should match the complexity of the function underlying the data. If the hypothesis is less complex than the function, then the model has underfit the data. If the complexity of the model is increased in response, then the training error decreases. But if the hypothesis is too complex, then the model is subject to overfitting and generalization will be poorer.[26]

In addition to performance bounds, computational learning theorists study the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be done in polynomial time. There are two kinds of time complexity results. Positive results show that a certain class of functions can be learned in polynomial time. Negative results show that certain classes cannot be learned in polynomial time.

Approaches

Decision tree learning

Decision tree learning uses a decision tree as a predictive model, which maps observations about an item to conclusions about the item's target value.

Association rule learning

Association rule learning is a method for discovering interesting relations between variables in large databases.

Artificial neural networks

An artificial neural network (ANN) learning algorithm, usually called "neural network" (NN), is a learning algorithm that is inspired by the structure and functional aspects of biological neural networks. Computations are structured in terms of an interconnected group of artificial neurons, processing information using a connectionist approach to computation. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs, to find patterns in data, or to capture the statistical structure in an unknown joint probability distribution between observed variables.

Deep learning

Falling hardware prices and the development of GPUs for personal use in the last few years have contributed to the development of the concept of deep learning which consists of multiple hidden layers in an artificial neural network. This approach tries to model the way the human brain processes light and sound into vision and hearing. Some successful applications of deep learning are computer vision and speech recognition.[27]

Inductive logic programming

Inductive logic programming (ILP) is an approach to rule learning using logic programming as a uniform representation for input examples, background knowledge, and hypotheses. Given an encoding of the known background knowledge and a set of examples represented as a logical database of facts, an ILP system will derive a hypothesized logic program that entails all positive and no negative examples. Inductive programming is a related field that considers any kind of programming languages for representing hypotheses (and not only logic programming), such as functional programs.

Support vector machines

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

Clustering

Cluster analysis is the assignment of a set of observations into subsets (called *clusters*) so that observations within the same cluster are similar according to some predesignated criterion or criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some *similarity metric* and evaluated for example by *internal compactness* (similarity between members of the same cluster) and *separation* between different clusters. Other methods are based on *estimated density* and *graph connectivity*. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis.

Bayesian networks

A Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that represents a set of random variables and their conditional independencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. Efficient algorithms exist that perform inference and learning.

Reinforcement learning

Reinforcement learning is concerned with how an *agent* ought to take *actions* in an *environment* so as to maximize some notion of long-term *reward*. Reinforcement learning algorithms attempt to find a *policy* that maps *states* of the world to the actions the agent ought to take in those states. Reinforcement learning differs from the supervised learning problem in that correct input/output pairs are never presented, nor sub-optimal actions explicitly corrected.

Representation learning

Several learning algorithms, mostly unsupervised learning algorithms, aim at discovering better representations of the inputs provided during training. Classical examples include principal components analysis and cluster analysis. Representation learning algorithms often attempt to preserve the information in their input but transform it in a way that makes it useful, often as a pre-processing step before performing classification or predictions, allowing reconstruction of the inputs coming from the unknown data generating distribution, while not being necessarily faithful for configurations that are implausible under that distribution.

Manifold learning algorithms attempt to do so under the constraint that the learned representation is low-dimensional. Sparse coding algorithms attempt to do so under the constraint that the learned representation is sparse (has many zeros). Multilinear subspace learning algorithms aim to learn low-dimensional representations directly from tensor representations for multidimensional data, without reshaping them into (high-dimensional) vectors.[28] Deep learning algorithms discover multiple levels of representation, or a hierarchy of features, with higher-level, more abstract features defined in terms of (or generating) lower-level features. It has been argued that an intelligent machine is one that learns a representation that disentangles the underlying factors of variation that explain the observed data.[29]

Similarity and metric learning

In this problem, the learning machine is given pairs of examples that are considered similar and pairs of less similar objects. It then needs to learn a similarity function (or a distance metric function) that can predict if new objects are similar. It is sometimes used in Recommendation systems.

Sparse dictionary learning

In this method, a datum is represented as a linear combination of basis functions, and the coefficients are assumed to be sparse. Let x be a d-dimensional datum, D be a d by n matrix, where each column of D represents a basis function. r is the coefficient to represent x using D. Mathematically, sparse dictionary learning means solving $x \approx Dr$ where r is sparse. Generally speaking, n is assumed to be larger than d to allow the freedom for a sparse representation.

Learning a dictionary along with sparse representations is strongly NP-hard and also difficult to solve approximately.[30] A popular heuristic method for sparse dictionary learning is K-SVD.

Sparse dictionary learning has been applied in several contexts. In classification, the problem is to determine which classes a previously unseen datum belongs to. Suppose a dictionary for each class has already been built. Then a new datum is associated with the class such that it's best sparsely represented by the corresponding dictionary. Sparse dictionary learning has also been applied in image de-noising. The key idea is that a clean image patch can be sparsely represented by an image dictionary, but the noise cannot.[31]

Genetic algorithms

A genetic algorithm (GA) is a search heuristic that mimics the process of natural selection, and uses methods such as mutation and crossover to generate new genotype in the hope of finding good solutions to a given problem. In machine learning, genetic algorithms found some uses in the 1980s and 1990s.[32][33] Vice versa, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms.[34]

Rule-based machine learning

Rule-based machine learning is a general term for any machine learning method that identifies, learns, or evolves `rules' to store, manipulate or apply, knowledge. The defining characteristic of a rule-based machine learner is the identification and utilization of a set of relational rules that collectively represent the knowledge captured by the system. This is in contrast to other machine learners that commonly identify a singular model that can be universally applied to any instance in order to make a prediction.[35] Rule-based machine learning approaches include learning classifier systems, association rule learning, and artificial immune systems.

Learning classifier systems

Learning classifier systems (LCS) are a family of rule-based machine learning algorithms that combine a discovery component (e.g. typically a genetic algorithm) with a learning component (performing either supervised learning, reinforcement learning, or unsupervised learning). They seek to identify a set of context-dependent rules that collectively store and apply knowledge in a piecewise manner in order to make predictions.[36]

Applications

Applications for machine learning include:

- Adaptive websites
- Affective computing
- Bioinformatics
- Brain-machine interfaces
- Cheminformatics
- Classifying DNA sequences
- Computational anatomy
- Computer vision, including object recognition
- Detecting credit card fraud
- Game playing
- Information retrieval
- Internet fraud detection
- Marketing
- Machine learning control
- Machine perception
- Medical diagnosis
- Economics
- Natural language processing
- Natural language understanding
- Optimization and metaheuristic
- Online advertising
- Recommender systems
- Robot locomotion
- Search engines
- Sentiment analysis (or opinion mining)
- Sequence mining
- Software engineering
- Speech and handwriting recognition
- Financial market analysis
- Structural health monitoring
- Syntactic pattern recognition
- User behavior analytics
- Translation[37]

In 2006, the online movie company Netflix held the first "Netflix Prize" competition to find a program to better predict user preferences and improve the accuracy on its existing Cinematch movie recommendation algorithm by at least 10%. A joint team made up of researchers from AT&T Labs-Research in collaboration with the teams Big Chaos and Pragmatic Theory built an ensemble model to win the Grand Prize in 2009 for $1 million.[38] Shortly after the prize was awarded, Netflix realized that viewers' ratings were not the best indicators of their viewing patterns ("everything is a recommendation") and they changed their recommendation engine accordingly.[39]

In 2012, co-founder of Sun Microsystems Vinod Khosla predicted that 80% of medical doctors jobs would be lost in the next two decades to automated machine learning medical diagnostic software.[40]

In 2014, it has been reported that a machine learning algorithm has been applied in Art History to study fine art paintings, and that it may have revealed previously unrecognized influences between artists.[41]

Model assessments

Classification machine learning models can be validated by accuracy estimation techniques like the Holdout method, which splits the data in a training and test set (conventionally 2/3 training set and 1/3 test set designation) and evaluates the performance of the training model on the test set. In comparison, the N-fold-cross-validation method randomly splits the data in k subsets where the k-1 instances of the data are used to train the model while the kth instance is used to test the predictive ability of the training model. In addition to the holdout and cross-validation methods, bootstrap, which samples n instances with replacement from the dataset, can be used to assess model accuracy.[42]

In addition to overall accuracy, investigators frequently report sensitivity and specificity (True Positive Rate: TPR and True Negative Rate: TNR, respectively) meaning True Positive Rate (TPR) and True Negative Rate (TNR) respectively. Similarly, investigators sometimes report the False Positive Rate (FPR) as well as the False Negative Rate (FNR). However, these rates are ratios that fail to reveal their numerators and denominators. The Total Operating Characteristic (TOC) is an effective method to express a model's diagnostic ability. TOC shows the numerators and denominators of the previously mentioned rates, thus TOC provides more information than the commonly used Receiver operating characteristic (ROC) and ROC's associated Area Under the Curve (AUC).

Ethics

Machine Learning poses a host of ethical questions. Systems which are trained on datasets collected with biases may exhibit these biases upon use, thus digitizing cultural prejudices.[43] Responsible collection of data thus is a critical part of machine learning.

Because language contains biases, machines trained on language corpora will necessarily also learn bias.[44]

See Machine ethics for additional information.

Software

Software suites containing a variety of machine learning algorithms include the following :

Free and open-source software

- CNTK
- Deeplearning4j
- dlib
- ELKI
- GNU Octave
- H2O
- Mahout
- Mallet
- mlpy
- MLPACK
- MOA (Massive Online Analysis)
- MXNet
- ND4J: ND arrays for Java
- NuPIC
- OpenAI Gym
- OpenAI Universe
- OpenNN
- Orange
- R
- scikit-learn
- Shogun
- SMILE (https://github.com/haifengl/smile)
- TensorFlow
- Torch
- Yooreeka
- Weka

Proprietary software with free and open-source editions

- KNIME
- RapidMiner

Proprietary software

- Amazon Machine Learning
- Angoss KnowledgeSTUDIO
- Ayasdi
- Google Prediction API
- IBM SPSS Modeler
- KXEN Modeler
- LIONsolver
- Mathematica
- MATLAB
- Microsoft Azure Machine Learning
- Neural Designer
- NeuroSolutions
- Oracle Data Mining
- RCASE
- SAS Enterprise Miner
- SequenceL
- Skymind
- Splunk
- STATISTICA Data Miner

Journals

- *Journal of Machine Learning Research*
- *Machine Learning*
- *Neural Computation*

Conferences

- Conference on Neural Information Processing Systems
- International Conference on Machine Learning

See also

- Automatic reasoning
- Big data
- Computational intelligence
- Computational neuroscience
- Data science
- Ethics of artificial intelligence
- Existential risk from advanced artificial intelligence
- Explanation-based learning
- Quantum machine learning
- Important publications in machine learning
- List of machine learning algorithms
- List of datasets for machine learning research

- Similarity learning
- Machine Learning Applications in Bioinformatics

References

1. Machine Learning and Optimization (https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf)
2. http://www.britannica.com/EBchecked/topic/1116194/machine-learning This tertiary source reuses information from other sources but does not name them.
3. Ron Kohavi; Foster Provost (1998). "Glossary of terms" (http://ai.stanford.edu/~ronnyk/glossary.html). *Machine Learning*. 30: 271–274.
4. Machine learning and pattern recognition "can be viewed as two facets of the same field."
5. Dickson, Ben. "Exploiting machine learning in cybersecurity" (http://social.techcrunch.com/2016/07/01/exploiting-machine-learning-in-cybersecurity/). *TechCrunch*. Retrieved 2017-05-23.
6. Wernick, Yang, Brankov, Yourganov and Strother, Machine Learning in Medical Imaging, *IEEE Signal Processing Magazine*, vol. 27, no. 4, July 2010, pp. 25–38
7. Mannila, Heikki (1996). *Data mining: machine learning, statistics, and databases*. Int'l Conf. Scientific and Statistical Database Management. IEEE Computer Society.
8. Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". *Computing Science and Statistics*. 29 (1): 3–9.
9. "Dark Reading" (http://www.darkreading.com/threat-intelligence/3-flavors-of-machine-learning--who-what-and-where/a/d-id/1324278).
10. "AI Business" (http://aibusiness.org/lightcybers-jason-matlof-explains-how-magna-detects-hackers-before-they-attack/).
11. "Machine Learning: What it is and why it matters" (http://www.sas.com/it_it/insights/analytics/machine-learning.html). *www.sas.com*. Retrieved 2016-03-29.
12. "Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage" (http://www.gartner.com/newsroom/id/3412017). Retrieved 2017-04-10.
13. "Why Machine Learning Models Often Fail to Learn: QuickTake Q&A" (https://www.bloomberg.com/news/articles/2016-11-10/why-machine-learning-models-often-fail-to-learn-quicktake-q-a). *Bloomberg.com*. 2016-11-10. Retrieved 2017-04-10.
14. Simonite, Tom. "Microsoft says its racist chatbot illustrates how AI isn't adaptable enough to help most businesses" (https://www.technologyreview.com/s/603944/microsoft-ai-isnt-yet-adaptable-enough-to-help-businesses/). *MIT Technology Review*. Retrieved 2017-04-10.
15. Mitchell, T. (1997). *Machine Learning*. McGraw Hill. p. 2. ISBN 0-07-042807-7.
16. Harnad, Stevan (2008), "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence" (http://eprints.ecs.soton.ac.uk/12954/), in Epstein, Robert; Peters, Grace, *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Kluwer
17. Russell, Stuart; Norvig, Peter (2003) [1995]. *Artificial Intelligence: A Modern Approach* (2nd ed.). Prentice Hall. ISBN 978-0137903955.
18. Sarle, Warren. "Neural Networks and statistical models". *CiteseerX*. CiteSeerX 10.1.1.27.699 (https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.27.699).
19. Langley, Pat (2011). "The changing science of machine learning". *Machine Learning*. 82 (3): 275–279. doi:10.1007/s10994-011-5242-y (https://doi.org/10.1007%2Fs10994-011-5242-y).
20. Le Roux, Nicolas; Bengio, Yoshua; Fitzgibbon, Andrew (2012). "Improving First and Second-Order Methods by Modeling Uncertainty". In Sra, Suvrit; Nowozin, Sebastian; Wright, Stephen J. *Optimization for Machine Learning*. MIT Press. p. 404.
21. MI Jordan (2014-09-10). "statistics and machine learning" (http://www.reddit.com/r/MachineLearning/comments/2fxi6v/ama_michael_i_jordan/ckelmtt?context=3). reddit. Retrieved 2014-10-01.
22. Cornell University Library. "Breiman : Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)" (http://projecteuclid.org/download/pdf_1/euclid.ss/1009213726). Retrieved 8 August 2015.
23. Gareth James; Daniela Witten; Trevor Hastie; Robert Tibshirani (2013). *An Introduction to Statistical Learning* (http://www-bcf.usc.edu/~gareth/ISL/). Springer. p. vii.
24. Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer, ISBN 0-387-31073-8
25. Mohri, Mehryar; Rostamizadeh, Afshin; Talwalkar, Ameet (2012). *Foundations of Machine Learning*. USA, Massachusetts: MIT Press. ISBN 9780262018258.
26. Alpaydin, Ethem (2010). *Introduction to Machine Learning* (https://mitpress.mit.edu/books/introduction-machine-learning). London: The MIT Press. ISBN 978-0-262-01243-0. Retrieved 4 February 2017.
27. Honglak Lee, Roger Grosse, Rajesh Ranganath, Andrew Y. Ng. "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.802&rep=rep1&type=pdf)" Proceedings of the 26th Annual International Conference on Machine Learning, 2009.
28. Lu, Haiping; Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data" (http://www.dsp.utoronto.ca/~haiping/Publication/SurveyMSL_PR2011.pdf) (PDF). *Pattern Recognition*. 44 (7): 1540–1551. doi:10.1016/j.patcog.2011.01.004 (https://doi.org/10.1016%2Fj.patcog.2011.01.004).
29. Yoshua Bengio (2009). *Learning Deep Architectures for AI* (https://books.google.com/books?id=cq5ewg7FniMC&pg=PA3). Now Publishers Inc. pp. 1–3. ISBN 978-1-60198-294-0.
30. A. M. Tillmann, "On the Computational Intractability of Exact and Approximate Dictionary Learning (http://dx.doi.org/10.1109/LSP.2014.2345761)", IEEE Signal Processing Letters 22(1), 2015: 45–49.
31. Aharon, M, M Elad, and A Bruckstein. 2006. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation." Signal Processing, IEEE Transactions on 54 (11): 4311–4322
32. Goldberg, David E.; Holland, John H. (1988). "Genetic algorithms and machine learning". *Machine Learning*. 3 (2): 95–99. doi:10.1007/bf00113892 (https://doi.org/10.1007%2Fbf00113892).
33. Michie, D.; Spiegelhalter, D. J.; Taylor, C. C. (1994). *Machine Learning, Neural and Statistical Classification*. Ellis Horwood.
34. Zhang, Jun; Zhan, Zhi-hui; Lin, Ying; Chen, Ni; Gong, Yue-jiao; Zhong, Jing-hui; Chung, Henry S.H.; Li, Yun; Shi, Yu-hui (2011). "Evolutionary Computation Meets Machine Learning: A Survey" (http://ieeexplore.ieee.org/iel5/10207/6052357/06052374.pdf?arnumber=6052374) (PDF). *Computational Intelligence Magazine*. IEEE. 6 (4): 68–75. doi:10.1109/mci.2011.942584 (https://doi.org/10.1109%2Fmci.2011.942584).
35. Bassel, George W.; Glaab, Enrico; Marquez, Julietta; Holdsworth, Michael J.; Bacardit, Jaume (2011-09-01). "Functional Network Construction in Arabidopsis Using Rule-Based Machine Learning on Large-Scale Data Sets" (http://www.plantcell.org/content/23/9/3101). *The Plant Cell*. 23 (9): 3101–3116. doi:10.1105/tpc.111.088153 (https://doi.org/10.1105%2Ftpc.111.088153). ISSN 1532-298X (https://www.worldcat.org/issn/1532-298X). PMC 3203449 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3203449). PMID 21896882 (https://www.ncbi.nlm.nih.gov/pubmed/21896882).

36. Urbanowicz, Ryan J.; Moore, Jason H. (2009-09-22). "Learning Classifier Systems: A Complete Introduction, Review, and Roadmap" (http://www.hindawi.com/archive/2009/736398/). *Journal of Artificial Evolution and Applications*. 2009: 1–25. doi:10.1155/2009/736398 (https://doi.org/10.1155%2F2009%2F736398). ISSN 1687-6229 (https://www.worldcat.org/issn/1687-6229).
37. "AI-based translation to soon reach human levels: industry officials" (http://english.yonhapnews.co.kr/news/2017/01/10/0200000000AEN20170110009700320.html?did=2106m). Yonhap news agency. Retrieved 4 Mar 2017.
38. "BelKor Home Page" (http://www2.research.att.com/~volinsky/netflix/) research.att.com
39. "The Netflix Tech Blog: Netflix Recommendations: Beyond the 5 stars (Part 1)" (http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html). Retrieved 8 August 2015.
40. Vonod Khosla (January 10, 2012). "Do We Need Doctors or Algorithms?" (https://techcrunch.com/2012/01/10/doctors-or-algorithms/). Tech Crunch.
41. When A Machine Learning Algorithm Studied Fine Art Paintings, It Saw Things Art Historians Had Never Noticed (https://medium.com/the-physics-arxiv-blog/when-a-machine-learning-algorithm-studied-fine-art-paintings-it-saw-things-art-historians-had-never-b8e4e7bf7d3e), *The Physics at ArXiv blog*
42. Kohavi, Ron (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection" (http://web.cs.iastate.edu/~jtian/cs573/Papers/Kohavi-IJCAI-95.pdf) (PDF). *International Joint Conference on Artificial Intelligence*.
43. Bostrom, Nick (2011). "The Ethics of Artificial Intelligence" (http://www.nickbostrom.com/ethics/artificial-intelligence.pdf) (PDF). Retrieved 11 April 2016.
44. [1] (https://freedom-to-tinker.com/2016/08/24/language-necessarily-contains-human-biases-and-so-will-machines-trained-on-language-corpora/)

Further reading

- Nils J. Nilsson, *Introduction to Machine Learning (http://ai.stanford.edu/people/nilsson/mlbook.html)*.
- Trevor Hastie, Robert Tibshirani and Jerome H. Friedman (2001). *The Elements of Statistical Learning (http://www-stat.stanford.edu/~tibs/ElemStatLearn/)*, Springer. ISBN 0-387-95284-5.
- Pedro Domingos (September 2015), The Master Algorithm, Basic Books, ISBN 978-0-465-06570-7
- Ian H. Witten and Eibe Frank (2011). *Data Mining: Practical machine learning tools and techniques* Morgan Kaufmann, 664pp., ISBN 978-0-12-374856-0.
- Ethem Alpaydin (2004). Introduction to Machine Learning, MIT Press, ISBN 978-0-262-01243-0.
- David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms (http://www.inference.phy.cam.ac.uk/mackay/itila/book.html)* Cambridge: Cambridge University Press, 2003. ISBN 0-521-64298-1
- Richard O. Duda, Peter E. Hart, David G. Stork (2001) *Pattern classification* (2nd edition), Wiley, New York, ISBN 0-471-05669-3.
- Christopher Bishop (1995). *Neural Networks for Pattern Recognition*, Oxford University Press. ISBN 0-19-853864-2.
- Stuart Russell & Peter Norvig, (2002). *Artificial Intelligence - A Modern Approach*. Prentice Hall, ISBN 0-136-04259-7.
- Ray Solomonoff, *An Inductive Inference Machine*, IRE Convention Record, Section on Information Theory, Part 2, pp., 56-62, 1957.
- Ray Solomonoff, "An Inductive Inference Machine (http://world.std.com/~rjs/indinf56.pdf)" A privately circulated report from the 1956 Dartmouth Summer Research Conference on AI.

External links

- International Machine Learning Society (http://machinelearning.org/)
- Popular online course by Andrew Ng, at Coursera (https://www.coursera.org/learn/machine-learning). It uses GNU Octave. The course is a free version of Stanford University's actual course taught by Ng, whose lectures are also available for free (https://see.stanford.edu/Course/CS229).
- mloss (https://mloss.org/) is an academic database of open-source machine learning software.