

## Review Article

# Review on Computing Machinery and Intelligence

Jyoti S. Kale, Manjusha P. Dhoke and Pavan K. Dhoke

SRTMU, Nanded, Maharashtra, India

\*Corresponding author

## ABSTRACT

### Keywords

Machine,  
Computer and  
learning

"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  $E$ ."<sup>[16]</sup> This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?". In Turing's proposal the various characteristics that could be possessed by a *thinking machine* and the various implications in constructing one are exposed.

## Introduction

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed.<sup>[1]</sup>

Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959 while at IBM<sup>[2]</sup>. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence,<sup>[3]</sup> machine learning explores the study and construction of algorithms that can learn from and make predictions on data<sup>[4]</sup> – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions,<sup>[5]:2</sup> 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,<sup>[6]</sup> optical character recognition (OCR),<sup>[7]</sup> 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,<sup>[8]</sup> where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.<sup>[5]: vii[9]</sup> Machine

learning can also be unsupervised<sup>[10]</sup> and be used to learn and establish baseline behavioral profiles for various entities<sup>[11]</sup> 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.<sup>[12]</sup>

According to the Gartner hype cycle of 2016, machine learning is at its peak of inflated expectations.<sup>[13]</sup> Effective machine learning is difficult because finding patterns is hard and often not enough training data is available; as a result, machine-learning programs often fail to deliver.<sup>[14][15]</sup>

### **Types of problems and tasks**

Machine learning tasks are typically classified into two broad categories, depending on whether there is a learning "signal" or "feedback" available to a learning system:

**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. As special cases, the input signal can be only partially available, or restricted to special feedback:

**Semi-supervised learning:** the computer is given only an incomplete training signal: a training set with some (often many) of the target outputs missing.

**Active learning:** the computer can only obtain training labels for a limited set of

instances (based on a budget), and also has to optimize its choice of objects to acquire labels for. When used interactively, these can be presented to the user for labeling.

**Reinforcement learning:** training data (in form of rewards and punishments) is given only as feedback to the programs actions in a dynamic environment, such as driving a vehicle or playing a game against an opponent.<sup>[5]:3</sup>

**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).

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:<sup>[5]:3</sup>

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.<sup>[18]</sup> Probabilistic reasoning was also employed, especially in automated medical diagnosis.

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. By 1980, expert systems had come to dominate AI, and statistics was out of favor.<sup>[20]</sup> 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. 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 back propagation.

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. 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.

### **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. He also suggested the term data science as a placeholder to call the overall field.

Leo Breiman distinguished two statistical modelling paradigms: data model and algorithmic model, 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*.<sup>[24]</sup>

### **Computational learning theory**

A core objective of a learner is to generalize from its experience. 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 under fit 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 over fitting and generalization will be poorer.

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.

### **List of machine learning algorithms**

#### **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.

#### **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 machine**

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 pre designated 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.<sup>[29]</sup>

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.

### **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 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. 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.

### **Genetic algorithm**

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.<sup>[33][34]</sup> Conversely, machine learning techniques have been used to improve the performance of genetic and evolutionary algorithms.

### **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. Rule-based machine learning approaches include learning classifier systems, association rule learning, and artificial immune systems.

### **Learning classifier system**

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.

### **Applications**

Applications for machine learning include:

Automated theorem proving

Adaptive websites<sup>[citation needed]</sup>  
Affective computing  
Bioinformatics  
Brain-machine interfaces  
Cheminformatics  
Classifying DNA sequences  
Computational anatomy  
Computer Networks  
Computer vision, including object recognition  
Detecting credit-card fraud  
General game playing  
Information retrieval  
Internet fraud detection  
Linguistics  
Marketing  
Machine learning control  
Machine perception  
Medical diagnosis  
Economics  
Insurance  
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  
Time series forecasting  
User behavior analytics  
Translation<sup>[42]</sup>

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.<sup>[43]</sup> 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. In 2010 The Wall Street Journal wrote about the firm Rebellion Research and their use of Machine Learning to predict the financial crisis.

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. 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.

### **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.

In addition to overall accuracy, investigators frequently report sensitivity and specificity

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 (algorithmic bias), thus digitizing cultural prejudices.<sup>[49]</sup> For example, using job hiring data from a firm with racist hiring policies may lead to a machine learning system duplicating the bias by scoring job applicants against similarity to previous successful applicants. Responsible collection of data and documentation of algorithmic rules used by a system thus is a critical part of machine learning.

Because language contains biases, machines trained on language *corpora* will necessarily also learn bias.

### References

1. Supposedly paraphrased from: Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. 3 (3). doi:10.1147/rd.33.0210. Confer Koza,

John R., Bennett, Forrest H., Andre, David; Keane, Martin A. (1996). *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. *Artificial Intelligence in Design '96*. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4\_9. Paraphrasing Arthur Samuel (1959), the question is: How can computers learn to solve problems without being explicitly programmed?

2. R. Kohavi and F. Provost, 1998. Glossary of terms," *Machine Learning*, 30(2-3): 271-274.
3. <http://www.britannica.com/EBchecked/topic/1116194/machine-learning> This tertiary source reuses information from other sources but does not name them.
4. Ron Kohavi; Foster Provost (1998). "Glossary of terms". *Machine Learning*. 30: 271–274.
5. Machine learning and pattern recognition "can be viewed as two facets of the same field."
6. Dickson, Ben. "Exploiting machine learning in cybersecurity". *TechCrunch*. Retrieved 2017-05-23.
7. Wernick, Yang, Brankov, Yourganov and Strother, Machine Learning in Medical Imaging, *IEEE Signal Processing Magazine*, 27(4): 25–38.
8. Mannila, Heikki (1996). *Data mining: machine learning, statistics, and databases*. *Int'l Conf. Scientific and Statistical Database Management*. *IEEE Computer Society*.
9. Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". *Computing Science and Statistics*. 29 (1): 3–9.
10. "Dark Reading".
11. "AI Business".

12. *"Machine Learning: What it is and why it matters". www.sas.com. Retrieved 2016-03-29.*
13. *"Gartner's 2016 Hype Cycle for Emerging Technologies Identifies Three Key Trends That Organizations Must Track to Gain Competitive Advantage". Retrieved 2017-04-10.*
14. *"Why Machine Learning Models Often Fail to Learn: QuickTake Q&A". Bloomberg.com.*
15. *Simonite, Tom. "Microsoft says its racist chatbot illustrates how AI isn't adaptable enough to help most businesses". MIT Technology Review. Retrieved 2017-04-10.*
16. *Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2.*