Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statist
seen data, and thus perform tasks without explicit instructions.[1] Recently, generative artificial neural networks hav
.[2][3]

Machine learning approaches have been applied to many fields including large language models, computer vision, sp
it is too costly to develop algorithms to perform the needed tasks.[4][5] ML is known in its application across busines
ll machine learning is statistically based, computational statistics is an important source of the field's methods.

Artificial intelligence

Machine learning as subfield of AI[22]

As a scientific endeavor, machine learning grew out of the quest for artificial intelligence (AI). In the early days of AI a
d in having machines learn from data. They attempted to approach the problem with various symbolic methods, as w
ceptrons and other models that were later found to be reinventions of the generalized linear models of statistics.[23
ated medical diagnosis.[24]: 488

However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine le
al problems of data acquisition and representation.[24]: 488 By 1980, expert systems had come to dominate AI, and
learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was
information retrieval.[24]: 708–710, 755 Neural networks research had been abandoned by AI and computer science
CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart, and Hinton. Their r
ropagation.[24]: 25

Machine learning (ML), reorganized and recognized as its own field, started to flourish in the 1990s. The field change
olvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, a
gic, and probability theory.[25]

Data mining

Machine learning and data mining often employ the same methods and overlap significantly, but while machine lear
e training data, data mining focuses on the discovery of (previously) unknown properties in the data (this is the analy
s many machine learning methods, but with different goals; on the other hand, machine learning also employs data
to improve learner accuracy. Much of the confusion between these two research communities (which do often have
xception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated w
nowledge discovery and data mining (KDD) the key task is the discovery of previously unknown knowledge. Evaluate
d will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot

Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of s
press the discrepancy between the predictions of the model being trained and the actual problem instances (for exa
and models are trained to correctly predict the pre-assigned labels of a set of examples).[26]

Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher"
ts.

Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its inpu
g 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 certa
ent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries t

Although each algorithm has advantages and limitations, no single algorithm works for all problems.[36][37][38]

Supervised learning

Main article: Supervised learning

A support-vector machine is a supervised learning model that divides the data into regions separated by a linear bou
he white.

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the de
s of a set of training examples. Each training example has one or more inputs and the desired output, also known as
ample is represented by an array or vector, sometimes called a feature vector, and the training data is represented b
ion, supervised learning algorithms learn a function that can be used to predict the output associated with new inpu
ermine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its
rform that task.[19]

Types of supervised-learning algorithms include active learning, classification and regression.[41] Classification algor
et of values, and regression algorithms are used when the outputs may have any numerical value within a range. As
input would be an incoming email, and the output would be the name of the folder in which to file the email.

Similarity learning is an area of supervised machine learning closely related to regression and classification, but the g
at measures how similar or related two objects are. It has applications in ranking, recommendation systems, visual i

Unsupervised learning
Main article: Unsupervised learning
See also: Cluster analysis
Unsupervised learning algorithms find structures in data that has not been labeled, classified or categorized. Insteac
entify commonalities in the data and react based on the presence or absence of such commonalities in each new pie
clude clustering, dimensionality reduction,[8] and density estimation.[42] Unsupervised learning algorithms also stre
of a gene of interest from pan-genome.[43]

Clustering via Large Indel Permuted Slopes, CLIPS, turns the alignment image into a learning regression problem. Th
bles to identify segments sharing the same set of indels.
Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within th
nated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make
ed by some similarity metric and evaluated, for example, by internal compactness, or the similarity between membe
ers. Other methods are based on estimated density and graph connectivity.

Semi-supervised learning
Main article: Semi-supervised learning
Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised lea
g examples are missing training labels, yet many machine-learning researchers have found that unlabeled data, whe
ce a considerable improvement in learning accuracy.

In weakly supervised learning, the training labels are noisy, limited, or imprecise; however, these labels are often che
44]

Reinforcement learning
Main article: Reinforcement learning
Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in
Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations re
ulti-agent systems, swarm intelligence, statistics and genetic algorithms. In reinforcement learning, the environment
ny reinforcements learning algorithms use dynamic programming techniques.[45] Reinforcement learning algorithm
are used when exact models are infeasible. Reinforcement learning algorithms are used in autonomous vehicles or

Dimensionality reduction
Dimensionality reduction is a process of reducing the number of random variables under consideration by obtaining
f reducing the dimension of the feature set, also called the "number of features". Most of the dimensionality reducti
extraction. One of the popular methods of dimensionality reduction is principal component analysis (PCA). PCA invo
e (e.g., 2D). This results in a smaller dimension of data (2D instead of 3D), while keeping all original variables in the m
s proposes that high-dimensional data sets lie along low-dimensional manifolds, and many dimensionality reduction
arning and manifold regularization.

Other types
Other approaches have been developed which do not fit neatly into this three-fold categorization, and sometimes m
e, topic modeling, meta-learning.[48]

Self-learning
Self-learning, as a machine learning paradigm was introduced in 1982 along with a neural network capable of self-lea
th no external rewards and no external teacher advice. The CAA self-learning algorithm computes, in a crossbar fash
nsequence situations. The system is driven by the interaction between cognition and emotion.[50] The self-learning a
iteration executes the following machine learning routine:

Manifold learning algorithms attempt to do so under the constraint that the learned representation is low-dimensio
that the learned representation is sparse, meaning that the mathematical model has many zeros. Multilinear subsp
directly from tensor representations for multidimensional data, without reshaping them into higher-dimensional ve
resentation, or a hierarchy of features, with higher-level, more abstract features defined in terms of (or generating) l
achine is one that learns a representation that disentangles the underlying factors of variation that explain the obse

Feature learning is motivated by the fact that machine learning tasks such as classification often require input that i
wever, real-world data such as images, video, and sensory data has not yielded attempts to algorithmically define sp
epresentations through examination, without relying on explicit algorithms.

Sparse dictionary learning
Main article: Sparse dictionary learning
Sparse dictionary learning is a feature learning method where a training example is represented as a linear combina
method is strongly NP-hard and difficult to solve approximately.[59] A popular heuristic method for sparse dictiona
s been applied in several contexts. In classification, the problem is to determine the class to which a previously unse
has already been built, a new training example is associated with the class that is best sparsely represented by the c
een applied in image de-noising. The key idea is that a clean image patch can be sparsely represented by an image d

Robot learning
Robot learning is inspired by a multitude of machine learning methods, starting from supervised learning, reinforce

An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Her
ow represents a connection from the output of one artificial neuron to the input of another.
Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biologica
arn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

An ANN is a model based on a collection of connected units or nodes called "artificial neurons", which loosely mode
apses in a biological brain, can transmit information, a "signal", from one artificial neuron to another. An artificial ne
additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between a
al neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neur
e a weight that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a conne
nal is only sent if the aggregate signal crosses that threshold. Typically, artificial neurons are aggregated into layers.
tions on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly a

The original goal of the ANN approach was to solve problems in the same way that a human brain would. However,
viations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, spe
playing board and video games and medical diagnosis.

Deep learning consists of multiple hidden layers in an artificial neural network. This approach tries to model the way
ing. Some successful applications of deep learning are computer vision and speech recognition.[75]

Decision trees
Main article: Decision tree learning

A decision tree showing survival probability of passengers on the Titanic
Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represente
epresented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining, and machi
screte set of values are called classification trees; in these tree structures, leaves represent class labels, and branche
ss labels. Decision trees where the target variable can take continuous values (typically real numbers) are called reg
d to visually and explicitly represent decisions and decision making. In data mining, a decision tree describes data, b
ion-making.

Support-vector machines
Main article: Support-vector machine
Support-vector machines (SVMs), also known as support-vector networks, are a set of related supervised learning m
ng examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that pre
ning algorithm is a non-probabilistic, binary, linear classifier, although methods such as Platt scaling exist to use SVM
erforming linear classification, SVMs can efficiently perform a non-linear classification using what is called the kerne
l feature spaces.

Regression analysis
Main article: Regression analysis
Applications
There are many applications for machine learning, including:

Agriculture
Anatomy
Adaptive website
Affective computing
Astronomy
Automated decision-making
Banking
Behaviorism
Bioinformatics
Brain–machine interfaces
Cheminformatics
Citizen Science
Climate Science
Computer networks
Computer vision
Credit-card fraud detection
Data quality
DNA sequence classification
Economics
Financial market analysis[83]
General game playing
Handwriting recognition
Healthcare
Information retrieval
Insurance
Internet fraud detection
Knowledge graph embedding
Linguistics
Machine learning control
Machine perception
Machine translation
Marketing
Medical diagnosis
Natural language processing
Natural language understanding
Online advertising
Optimization
Recommender systems
Robot locomotion
Search engine.