

5. Machine Learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data.

The process of machine learning is similar to that of data mining. Both systems search through data to look for patterns. However, instead of extracting data for human comprehension, machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data. Unsupervised algorithms can draw inferences from datasets.

Facebook's News Feed uses machine learning to personalize each member's feed. If a member frequently stops scrolling in order to read or "like" a particular friend's posts, the News Feed will start to show more of that friend's activity earlier in the feed. Behind the scenes, the software is simply using statistical analysis and predictive analytics to identify patterns in the user's data and use to patterns to populate the News Feed.

Machine learning is particularly attractive in several real life problem because of the following reasons:

- Some tasks cannot be defined well except by example
- Working environment of machines may not be known at design time
- Explicit knowledge encoding may be difficult and not available
- Environments change over time
- Biological systems learn

Recently, learning is widely used in a number of application areas including,

- Data mining and knowledge discovery
- Speech/image/video (pattern) recognition
- Adaptive control
- Autonomous vehicles/robots
- Decision support systems
- Bioinformatics
- WWW

Formally, 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 .



Thus a learning system is characterized by:

- task T
- experience E, and
- performance measure P

Examples:

Learning to drive a van

T: Drive on a public highway using vision sensors

P: Average distance traveled before an error (according to human observer)

E: Sequence of images and steering actions recorded during human driving.

Learning

Learning is the act of acquiring new, or modifying and reinforcing, existing knowledge, behaviors, skills, values, or preferences and may involve synthesizing different types of information. The ability to learn is possessed by humans, animals, plants and some machines. Progress over time tends to follow a learning curve. It does not happen all at once, but builds upon and is shaped by previous knowledge. To that end, learning may be viewed as a process, rather than a collection of factual and procedural knowledge. Learning produces changes in the organism and the changes produced are relatively permanent.

The operation of a neural network is determined by the values of the interconnection weights. There is no algorithm that determines how the weights should be assigned in order to solve specific problems. Hence, the weights are determined by a learning process

- **Learning by Analogy**

Analogy is a reasoning process of transferring information from a particular subject to another. In a narrower sense, analogy is an inference or an argument from one particular to another particular, as opposed to deduction and induction where at least one of the evidences or the conclusion is general.

Analogy plays a significant role in problem solving such as, decision making, perception, memory, creativity, emotion, explanation, and communication. It lies behind basic tasks such as the identification of places, objects and people, for example, in face perception and facial recognition systems.

- **Inductive Learning**

Inductive Learning is a powerful strategy for helping machine deepen their understanding of content and develop their inference and evidence-gathering skills. In an Inductive Learning, Machines examine, group, and label specific "bits" of information to find patterns. For



example, if given 20 specific weather-related terms (e.g., *rain*, *sleet*, *snow*, *hygrometer*, *rain gauge*, *thermometer*, *humid*, *dry*, *windy*, *cold*), machine might group the terms into an initial set of categories labeled *precipitation*, *weather instruments*, and *how weather feels*. Inductive Learning does not stop at categorization, however; it also instruct machine to use their labeled groups to develop a set of working hypotheses about the content to come. Then, during the learning, machine collect evidence to verify or refine each of their hypotheses.

– **Explanation based Learning**

Explanation-based learning (EBL) is a form of machine learning that exploits a very strong, or even perfect, domain theory to make generalizations or form concepts from training examples.

An example of EBL using a perfect domain theory is a program that learns to play chess by being shown examples. A specific chess position that contains an important feature, say, "Forced loss of black queen in two moves," includes many unrelated features, such as the specific scattering of pawns on the board. EBL can take a single training example and determine what the relevant features are in order to form a generalization

Learning may be classified into two categories:

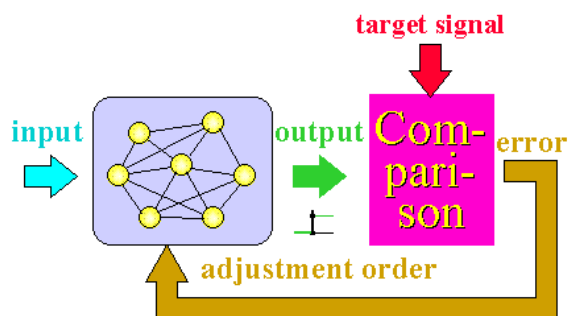
- a. **Supervised Learning**
- b. **Unsupervised Learning**

Consider a machine (or living organism) which receives some sequence of inputs X_1, X_2, X_3, \dots , where X_t is the sensory input at time t . In supervised learning the machine is given a sequence of input & a sequence of desired outputs y_1, y_2, \dots , and the goal of the machine is to learn to produce the correct output given a new input. While, in unsupervised learning the machine simply receives inputs X_1, X_2, X_3, \dots , but obtains neither supervised target outputs, nor rewards from its environment.

Supervised Learning:

In supervised learning, the network is presented with inputs together with the target (teacher signal) outputs. Then, the neural network tries to produce an output as close as possible to the target signal by adjusting the values of internal weights. The most common supervised learning method is the “error correction method”.

Error correction method is used for networks which their neurons have discrete output functions. Neural



networks are trained with this method in order to reduce the error (difference between the network's output and the desired output) to zero.

- **Naïve Bayes Classifier**

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.

- **Logistics Regression**

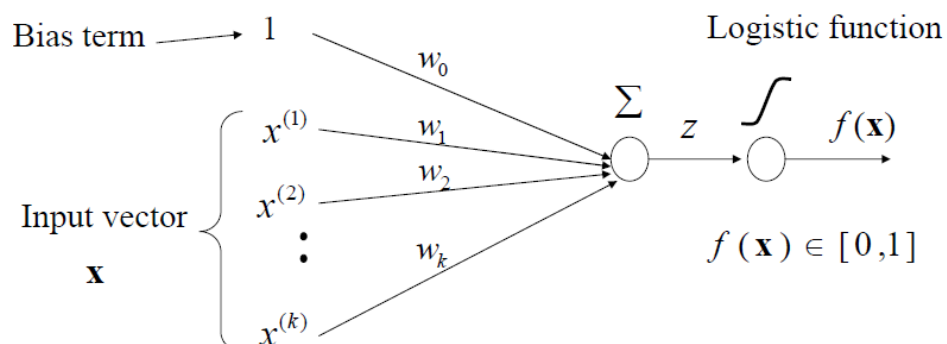
Logistic regression is used whenever you need to assign an input to one of several classes. It applies a logistic function to a linear combination of features, yielding a probability for an input to be in one of the classes. The output is usually binary, but logistic regression can also be applied to multiclass classification problems.

A function model with smooth switching:

$$f(\mathbf{x}) = g(w_0 + w_1x^{(1)} + \dots w_kx^{(k)})$$

where \mathbf{w} are parameters of the models

and $g(z)$ is a **logistic function** $g(z) = 1/(1 + e^{-z})$



- **Support Vector Machine**

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

Unsupervised Learning:



In unsupervised learning, there is no teacher (target signal) from outside and the network adjusts its weights in response to only the input patterns.

- **Clustering**

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, bioinformatics and data compression.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem.

- **K-means clustering**

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space.

- **Dimensionality Reduction**

In machine learning and statistics, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration, via obtaining a set "uncorrelated" principle variables. It can be divided into feature selection and feature extraction.

For high-dimensional datasets (i.e. with number of dimensions more than 10), dimension reduction is usually performed prior to applying a K-nearest neighbors algorithm (k-NN) in order to avoid the effects of the curse of dimensionality.

Reinforcement Learning

In reinforcement learning a teacher is available, but the teacher instead of directly providing the desired action corresponding to a perception, return reward and punishment to the learner for its action corresponding to a perception. Examples include a robot in an unknown topography where it get a punishment when its hits an obstacle and reward when it moves smoothly. In order to design a learning system the designer has to make the following choices based on the application.

Fuzzy Logic

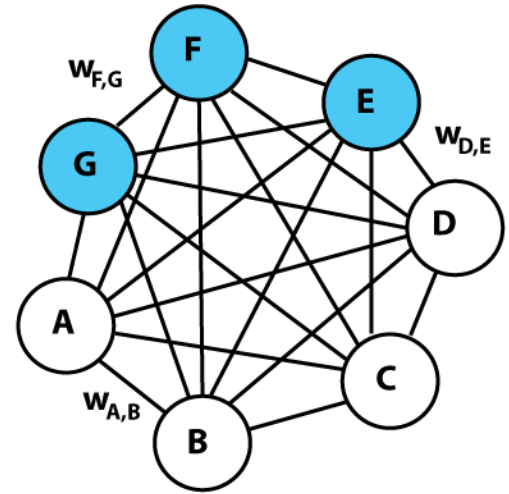
Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number



between 0 and 1, considered to be "fuzzy"/"Uncertain". By contrast, in Boolean logic, the truth values of variables may only be 0 or 1. Fuzzy logic has been employed to handle the concept of partial truth, where the truth value may range between completely true and completely false. Furthermore, when linguistic variables are used, these degrees may be managed by specific (membership) functions.

Boltzmann Machines

A Boltzmann machine is a type of neural network. Boltzmann machines can be seen as the stochastic (being or having a random variable.), multiplicative equivalent of Hopfield nets. They were one of the first examples of a neural network capable of learning internal representations, and are able to represent and (given sufficient time) solve difficult combinatorial problems. They are theoretically stimulating because of the locality and Hebbian nature of their training algorithm, and because of their parallelism and the resemblance of their dynamics to simple physical processes.



A Boltzmann machine, like a Hopfield network, is a network of units with an "energy" defined for the network. It also has binary units, but unlike Hopfield nets, Boltzmann machine units are stochastic. The global energy, E , in a Boltzmann machine is identical in form to that of a Hopfield network:

$$E = - \left(\sum_{i < j} w_{ij} s_i s_j + \sum_i \theta_i s_i \right)$$

Where:

w_{ij} is the connection strength between unit j and unit i .

s_i is the state, $s_i \in \{0, 1\}$, of unit i .

θ_i is the bias of unit i in the global energy function. ($-\theta_i$ is the activation threshold for the unit.)

The connections in a Boltzmann machine have two restrictions:

$w_{ii} = 0 \quad \forall i$. (No unit has a connection with itself.)

$w_{ij} = w_{ji} \quad \forall i, j$. (All connections are symmetric.)

Often the weights are represented in matrix form with a symmetric matrix W , with zeros along the diagonal

Deep Learning



Deep learning (deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers, with complex structures or otherwise, composed of multiple non-linear transformations.

Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc. Some representations are better than others at simplifying the learning task (e.g., face recognition or facial expression recognition) from examples. One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction

Genetic Algorithms

- Genetic Algorithms are search algorithms based on the mechanism of natural selection and natural genetics.
- A genetic algorithm is a directed random search procedure.
- A genetic algorithm borrows ideas from biology to search a solution for a target value.
- GA process may include

