

Overview of Machine Learning

Machine learning is a field of data analysis that involves training complex algorithms for use in pattern recognition. In this definition, data includes any information or values that can be analyzed by these machine learning algorithms. Data must be collected and organized into a usable format so that it may be fed into a computer system, and the system will then recognize patterns in the data and make decisions based on those patterns. Having learned certain patterns from existing data, the system can then make predictions on future data. Generally, the designers of the system will train a machine learning algorithm on a contained set of data and evaluate the accuracy using a unique, unseen set of test data. Target accuracy for a machine learning system differs depending on the subject material and context, but a rule of thumb is that the algorithm should be significantly better at predicting trends than random guessing.

Artificial intelligence is a subset of computer science that involves trying to simulate human intelligence with processes like learning and problem solving. It is closely related to machine learning in that most instances of developing artificial intelligence require the ability to learn from data, which is the crux of machine learning.

In the modern world, machine learning is frequently used by big tech conglomerates in place of traditional programming because the knowledge required to perform certain tasks is not known beforehand and must be deduced from data. Google's search engine autofills searches based on common searches from millions of users; this could not be done with traditional algorithms because there are far too many possible combinations of words in a given search to manually fill out without automated data collection and pattern applications. Targeted advertising also uses advanced machine learning techniques; algorithms use searches, messages, and purchase histories to predict the types of products a user may be interested in for advertising purposes. The connections between all of this personal data and a specific advertisement are often far too unclear to be made with traditional search algorithms or manual presets, so machine learning systems process the volumes of data and generate predictions instead.

In its simplest and clearest form, data captured by machine learning systems consists of features and observations. An observation, sometimes called an instance, is a single data point of associated information. A feature is an individual attribute of data, like a column or category of information about a set of data points. Each feature can be quantitative, supporting a continuous numerical range, or qualitative, including only descriptive and discrete information. For supervised machine learning, the feature that requires predicting is the target variable, whereas the other features are predictors. Training data consisting of many observations is then passed into the system, where complex statistical algorithms find patterns between the predictors and the desired target for each observation.

Personally, I am curious about machine learning due to its status as a shiny new technology with far-reaching applications. I do not have enough work experience to pinpoint which subset of computer science I want to work in, but regardless of whichever field that may be, I am likely to come across machine learning and artificial intelligence. Many of the current

applications require vast data collection and surveillance that tends to lie in an ethical gray area. A personal project relevant to some of my current hobbies could be using machine learning for rating algorithms; existing algorithms like Elo and TrueSkill are used for chess and sports, among other competitive ventures, and they tend to be lacking in terms of predictive accuracy. While I definitely will not outdo top researchers for a generic win rate algorithm, I could try and see what specific parameters increase the predictive power for a more specialized ranking system.