Methods of Machine Learning

Types of Machine Learning Algorithms

The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve.

Supervised Learning: Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs.[27] The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and the desired output, also known as a supervisory signal. In the mathematical model, each training example is represented by an array or vector, sometimes called a feature vector, and the training data is represented by a matrix. Through iterative optimization of an , supervised learning algorithms learn a function that can be used to predict the output associated with new inputs.[28] An optimal function will allow the algorithm to correctly determine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its outputs or predictions over time is said to have learned to perform that task.

Types of supervised algorithms include Classification and Regression. Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. For example in classification problems classifying whether an email is spam or not.

Unsupervised Learning: Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labelled , classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. A central application of unsupervised learning is in the field of density estimation in statistics, such as finding the probability density function. Though unsupervised learning encompasses other domains involving summarizing and explaining data features.

Reinforcement Learning: When you present the algorithm with examples that lack labels, as in unsupervised learning. However, you can accompany an example with positive or negative feedback according to the solution the algorithm proposes comes under the category of Reinforcement learning, which is connected to applications for which the algorithm must make decisions (so the product is prescriptive, not just descriptive, as in unsupervised learning), and the decisions bear consequences. In the human world, it is just like learning by trial and error. In this case, an application presents the algorithm with examples of specific situations, such as having the gamer stuck in a maze while avoiding an enemy. The application lets the algorithm know the outcome of actions it takes, and learning occurs while trying to avoid what it discovers to be dangerous and to pursue survival.

Semi-supervised Learning: where an incomplete training signal is given: a training set with some (often many) of the target outputs missing. There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing.

Categorizing on the basis of required Output

Classification: When 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”.

Regression: Which is also a supervised problem, A case when the outputs are continuous rather than discrete.

Clustering: When 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.

Modeling in Machine Learning

Model Selection: We can think of the process of configuring and training the model as a model selection process. Each iteration we have a new model that we could choose to use or to modify. Even the choice of machine learning algorithm is part of that model selection process. Of all the possible models that exist for a problem, a given algorithm and algorithm configuration on the chosen training dataset will provide a finally selected model.

Inductive Bias: Bias is the limits imposed on the selected model. All models are biased which introduces error in the model, and by definition all models have error (they are generalizations from observations). Biases are introduced by the generalizations made in the model including the configuration of the model and the selection of the algorithm to generate the model. A machine learning method can create a model with a low or a high bias and tactics can be used to reduce the bias of a highly biased model.

Model Variance: Variance is how sensitive the model is to the data on which it was trained. A machine learning method can have a high or a low variance when creating a model on a dataset. A tactic to reduce the variance of a model is to run it multiple times on a dataset with different initial conditions and take the average accuracy as the model’s performance.

Bias-Variance Tradeoff: Model selection can be thought of as a trade-off of the bias and variance. A low bias model will have a high variance and will need to be trained for a long time or many times to get a usable model. A high bias model will have a low variance and will train quickly, but suffer poor and limited performance.