

# Some Key Machine Learning Definitions



joydeep bhattacharjee [Follow](#)

Oct 28, 2017 · 3 min read



**Model:** A machine learning model can be a mathematical representation of a real-world process. To generate a machine learning model you will need to provide training data to a machine learning algorithm to learn from.

**Algorithm:** Machine Learning algorithm is the hypothesis set that is taken at the beginning before the training starts with real-world data. When we say Linear Regression algorithm, it means a set of functions that define similar characteristics as defined by Linear Regression and from those set of functions we will choose one function that fits the most by the training data.

**Training:** While training for machine learning, you pass an algorithm with training data. The learning algorithm finds patterns in the training data such that the input parameters correspond to the target. The output of the training process is a machine learning model which you can then use to make predictions. This process is also called “learning”.

**Regression:** Regression techniques are used when the output is real-valued based on continuous variables. For example, any time series data. This technique involves fitting a line.

**Classification:** In classification, you will need to categorize data into predefined classes. For example, an email can either be ‘spam’ or ‘not spam’.

**Target:** The target is whatever the output of the input variables. It could be the individual classes that the input variables maybe mapped to in case of a classification problem or the output value range in a regression problem. If the training set is considered then the target is the training output values that will be considered.

**Feature:** Features are individual independent variables that act as the input in your system. Prediction models use features to make predictions. New features can also be obtained from old features using a method known as ‘feature engineering’. More simply, you can consider one column of your data set to be one feature. Sometimes these are also called **attributes**. And the number of features are called **dimensions**.

**Label:** Labels are the final output. You can also consider the output classes to be the labels. When data scientists speak of labeled data, they mean groups of samples that have been tagged to one or more labels.

**Overfitting:** An important consideration in machine learning is how well the approximation of the target function that has been trained using training data, generalizes to new data. Generalization works best if the signal or the sample that is used as the training data has a high signal to noise ratio. If that is not the case, generalization would be poor and we will not get good predictions. A model is overfitting if it fits the training data too well and there is a poor generalization of new data.

**Regularization:** Regularization is the method to estimate a preferred complexity of the machine learning model so that the model generalizes and the over-fit/under-fit problem is avoided. This is done by adding a penalty on the different parameters of the model thereby reducing the freedom of the model.

**Parameter and Hyper-Parameter:** Parameters are configuration variables that can be thought to be internal to the model as they can be

estimated from the training data. Algorithms have mechanisms to optimize parameters. On the other hand, hyperparameters cannot be estimated from the training data. Hyperparameters of a model are set and tuned depending on a combination of some heuristics and the experience and domain knowledge of the data scientist.

These are some key machine learning terms that I thought are important and should be looked into for someone who is a machine learning beginner. In case you have encountered some common terms which are not included here, do write about them in the comments below.

Thanks for reading this. If you are interested in talking more on this, just drop me a message [@alt227Joydeep](#). I would be glad to discuss this further. Also please hit the claps and help this article reach more audience.

Reference:

- [Difference between machine learning model and algorithm.](#)
- [Difference between labeled and unlabeled data.](#)
- [label and feature.](#)
- [what is a feature.](#)
- [overfitting and underfitting](#)
- [dimension, attributes and features.](#)
- [parameter and hyper parameter.](#)
- [stats.stackexchange. regularization.](#)
- [regularisation.explained](#)

