

[Open in app](#)

Published in Towards Data Science



Sidath Asiri

[Follow](#)

Jun 11, 2018 · 7 min read



Save



# Machine Learning Classifiers



## What is classification?

Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function ( $f$ ) from input variables ( $X$ ) to discrete output variables ( $y$ ).

For example, spam detection in email service providers can be identified as a



to be used as the training data. When the classifier is trained accurately, it can be used to detect an unknown email.

Classification belongs to the category of supervised learning where the targets also provided with the input data. There are many applications in classification in many domains such as in credit approval, medical diagnosis, target marketing etc.

There are two types of learners in classification as lazy learners and eager learners.

### 1. Lazy learners

Lazy learners simply store the training data and wait until a testing data appear. When it does, classification is conducted based on the most related data in the stored training data. Compared to eager learners, lazy learners have less training time but more time in predicting.

*Ex. k-nearest neighbor, Case-based reasoning*

### 2. Eager learners

Eager learners construct a classification model based on the given training data before receiving data for classification. It must be able to commit to a single hypothesis that covers the entire instance space. Due to the model construction, eager learners take a long time for train and less time to predict.

*Ex. Decision Tree, Naive Bayes, Artificial Neural Networks*

## Classification algorithms

There is a lot of classification algorithms available now but it is not possible to conclude which one is superior to other. It depends on the application and nature of available data set. For example, if the classes are linearly separable, the linear classifiers like Logistic regression, Fisher's linear discriminant can outperform sophisticated models and vice versa.

### Decision Tree



Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time. Each time a rule is learned, the tuples covered by the rules are removed. This process is continued on the training set until meeting a termination condition.

The tree is constructed in a top-down recursive divide-and-conquer manner. All the attributes should be categorical. Otherwise, they should be discretized in advance. Attributes in the top of the tree have more impact towards in the classification and they are identified using the information gain concept.

A decision tree can be easily over-fitted generating too many branches and may reflect anomalies due to noise or outliers. An over-fitted model has a very poor performance on the unseen data even though it gives an impressive performance on training data. This can be avoided by pre-pruning which halts tree construction early or post-pruning which removes branches from the fully grown tree.

### **Naive Bayes**

Naive Bayes is a probabilistic classifier inspired by the Bayes theorem under a simple assumption which is the attributes are conditionally independent.



The classification is conducted by deriving the maximum posterior which is the maximal  $P(C_i | \mathbf{X})$  with the above assumption applying to Bayes theorem. This assumption greatly reduces the computational cost by only counting the class distribution. Even though the assumption is not valid in most cases since the attributes are dependent, surprisingly Naive Bayes has able to perform impressively.

Naive Bayes is a very simple algorithm to implement and good results have obtained in most cases. It can be easily scalable to larger datasets since it takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

Naive Bayes can suffer from a problem called the zero probability problem. When the conditional probability is zero for a particular attribute, it fails to give a valid prediction. This needs to be fixed explicitly using a Laplacian estimator.

## Artificial Neural Networks





to develop and test computational analogs of neurons. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples.

There are many network architectures available now like Feed-forward, Convolutional, Recurrent etc. The appropriate architecture depends on the application of the model. For most cases feed-forward models give reasonably accurate results and especially for image processing applications, convolutional networks perform better.

There can be multiple hidden layers in the model depending on the complexity of the function which is going to be mapped by the model. Having more hidden layers will enable to model complex relationships such as deep neural networks.

However, when there are many hidden layers, it takes a lot of time to train and adjust weights. The other disadvantage of is the poor interpretability of model compared to other models like Decision Trees due to the unknown symbolic meaning behind the learned weights.

But Artificial Neural Networks have performed impressively in most of the real world applications. It is high tolerance to noisy data and able to classify untrained patterns. Usually, Artificial Neural Networks perform better with continuous-valued inputs and outputs.

---

*All of the above algorithms are eager learners since they train a model in advance to generalize the training data and use it for prediction later.*

---

### **k-Nearest Neighbor (KNN)**



*k*-Nearest Neighbor is a lazy learning algorithm which stores all instances correspond to training data points in *n*-dimensional space. When an unknown discrete data is received, it analyzes the closest *k* number of instances saved (nearest neighbors) and returns the most common class as the prediction and for real-valued data it returns the mean of *k* nearest neighbors.

In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the *k* neighbors according to their distance using the following query giving greater weight to the closest neighbors.



Distance calculating query

Usually KNN is robust to noisy data since it is averaging the *k*-nearest neighbors.

[Open in app](#)

After training the model the most important part is to evaluate the classifier to verify its applicability.

### Holdout method

There are several methods exists and the most common method is the holdout method. In this method, the given data set is divided into 2 partitions as test and train 20% and 80% respectively. The train set will be used to train the model and the unseen test data will be used to test its predictive power.

### Cross-validation



--



12



Over-fitting is a common problem in machine learning which can occur in most models. k-fold cross-validation can be conducted to verify that the model is not over-fitted. In this method, the data-set is randomly partitioned into  $k$  *mutually exclusive* subsets, each approximately equal size and one is kept for testing while others are used for training. This process is iterated throughout the whole  $k$  folds.

### Precision and Recall

Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. Precision and Recall are used as a measurement of the relevance.

[Open in app](#)

ROC curve is a measure of the accuracy of the model. When a model is closer to the diagonal, it is less accurate and the model with perfect accuracy will have an area of 1.0



ROC Curve

[About](#) [Help](#) [Terms](#) [Privacy](#)

Get the Medium app



A button that says 'Download on the App Store', and if clicked it will lead you to the iOS App store



A button that says 'Get it on, Google Play', and if clicked it will lead you to the Google Play store



