# Machine learning Algorithms

## Linear regression

Finds coefficients  
Remove similar / correlated variables to remove noise from your data.

## Logistic Regression

Good for binary classification  
Transform output using the logistic function  
Remove similar / correlated variables to remove noise from your data.

## Linear Discriminant Analysis

Linear classification for more than two classes.  
Using statistical properties calculated for each class (such as mean and variance)  
Assumes Gaussian distribution, so removing outliers of data beforehand is smart.

## Classification and Regression Trees

Decision Trees is simply walking from the root node down to a leaf node for classification.  
Trees are fast to learn and very fast for making predictions. They are also often accurate for a broad range of problems and do not require any special preparation for your data.

## Naive Bayes

Probability of each class, The conditional probability for each class given each x value.  
Naive because it assumes that each input variable is independent. Technique is very effective on a large range of complex problems.

## K-Nearest Neighbors

Model is the entire training set. Predictions are made by looking at K most similar cases and summarizing the output variable for those. For regression this is the mean of outputs and for classification is could be the most common.

## Learning Vector Quantization

Summarize the KNN at random to a collection of codebook vectors. Predict like KNN but with most similar neighbor or best matching codebook vector. Uses ANN algorithm to summarize the KNN. Should be rescaled to similar range (0,1).

## Support Vector Machines

Probably most popular and talked about machine learning algorithms. A hyperplane that separates the points in the input variable space by their class. SVM learning algorithm find the coefficients that results in the best separation of the classes by the hyperplane.

## Bagging and Random Forest

Type of ensemble machine learning algorithm called Bootstrap Aggregation or bagging. Bootstrap is estimating a quantity from a data sample. Take lots of samples of your data, calculate the mean, then average all of your mean values to give a better estimation of the true mean value. Bagging is the same type of approach. Models are constructed for each data sample. When making a prediction each models make a prediction and the predictions are averaged to give a better estimate of the true output value.

Random forest is a tweak on this approach where decision trees are created so that rather than selecting optimal split points, suboptimal splits are made by introducing randomness. This way the models are more different. Combining their predictions results in a better estimate of the true underlying output value.

If you get good results with an algorithm with high variance (decision trees), you can often get better results by bagging that algorithm.

## Boosting and AdaBoost

Boosting is an ensemble technique that attempts to create a strong classifier from a number of weak classifiers. In essence, build a model from the training data, then create a second model that attempts to correct the first model. You do this until the training set is predicted perfectly or maximum number of models are added.

AdaBoost is a boosting algorithm developed for binary classification. And most modern boosting methods build on AdaBoost, most notably stochastic gradient boosting machines. (sequential)

## Ensemble methods

Good because they

1. average out biases.   
   If you average a bunch of democratic-leaning polls and republican-leaning polls together, you will get an average something that isn’t leaning either way.
2. They reduce variance.   
   The aggregate opinion of a bunch of models is less noisy than the single opinion of one of the models. In finance, this is called diversification — a mixed portfolio of many stocks will be much less variable than just one of the stocks alone. This is why your models will be better with more data points rather than fewer.
3. Unlikely to over-fit.  
   If you have individual models that didn’t over-fit, and you are combining the predictions from each model in a simple way (average, weighted average, logistic regression), then there’s no room for over-fitting.

## PCA – Principal component analysis

Is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

## Clustering Algorithms

Grouping a set of objects such that objects in the same group are more similar to each other than to those in other groups. There exists several algorithms and they work different.

## Singular Value Decomposition

Is a factorization of a real complex matrix. PCA is actually a simple application of SVD.

## Independent Component Analysis

ICA is a statistical technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non-gaussian and mutually independent, and they are called independent components of the observed data.

ICA is related to PCA, but more powerful because it is capable of finding the underlying factors of sources when these classic methods fail completely. Most common applications include digital images, document databases, economic indicators and psychometric measurements.

## In short:

The algorithm to use is depending on many factors such as size, quality and nature of data, the available computational time and the urgency of the task and lastly what you want to do with the data.