# Lesson 1: How To Talk About Data in Machine Learning

It is important to understand and use the right terminology when talking about data.  
  
How do you think about data? Think of a spreadsheet. You have columns, rows, and cells.  
  
The statistical perspective of machine learning frames data in the context of a hypothetical function (f) that the machine learning algorithm aims to learn. Given some input variables (Input)  the function answer the question as to what is the predicted output variable (Output).  
  
Output = f(Input)  
  
The inputs and outputs can be referred to as variables or vectors.  
  
The computer science perspective uses a row of data to describe an entity (like a person) or an observation about an entity. As such, the columns for a row are often referred to as attributes of the observation and the rows themselves are called instances.  
  
In the next lesson you will discover the principle that underpins all machine learning algorithms.

# The Principle That Underpins All Algorithms

Hi, there is a common principle that underlies all supervised machine learning algorithms for predictive modeling.

Machine learning algorithms are described as learning a target function (f) that best maps input variables (X) to an output variable (Y).

Y = f(X)

This is a general learning task where we would like to make predictions in the future (Y) given new examples of input variables (X). We don't know what the function (f) looks like or it's form. If we did, we would use it directly and we would not need to learn it from data using machine learning algorithms.

The most common type of machine learning is to learn the mapping Y = f(X) to make predictions of Y for new X. This is called predictive modeling or predictive analytics and our goal is to make the most accurate predictions possible.  
  
In the next lesson you will discover the difference between parametric and nonparametric algorithms.

# Lesson 3: Parametric and Nonparametric Algorithms

Hi, what is a parametric machine learning algorithm and how is it different from a nonparametric machine learning algorithm?

Assumptions can greatly simplify the learning process, but can also limit what can be learned. Algorithms that simplify the function to a known form are called parametric machine learning algorithms.

The algorithms involve two steps:

1. Select a form for the function.
2. Learn the coefficients for the function from the training data.

Some examples of parametric machine learning algorithms are Linear Regression and Logistic Regression.

Algorithms that do not make strong assumptions about the form of the mapping function are called nonparametric machine learning algorithms. By not making assumptions, they are free to learn any functional form from the training data.

Non-parametric methods are often more flexible, achieve better accuracy but require a lot more data and training time.

Examples of nonparametric algorithms include Support Vector Machines, Neural Networks and Decision Trees.  
  
In the next lesson you will discover the bias-variance trade-off.

# Lesson 4: Bias, Variance and the Trade-off

Bias are the simplifying assumptions made by a model to make the target function easier to learn.

Generally parametric algorithms have a high bias making them fast to learn and easier to understand but generally less flexible. In turn they have lower predictive performance on complex problems that fail to meet the simplifying assumptions of the algorithms bias.

Decision trees are an example of a low bias algorithm, whereas linear regression is an example of a high-bias algorithm.

Variance is the amount that the estimate of the target function will change if different training data was used. The target function is estimated from the training data by a machine learning algorithm, so we should expect the algorithm to have some variance, not zero variance.

The k-Nearest Neighbors algorithm is an example of a high-variance algorithm, whereas Linear Discriminant Analysis is an example of a low variance algorithm.

The goal of any predictive modeling machine learning algorithm is to achieve low bias and low variance. In turn the algorithm should achieve good prediction performance. The parameterization of machine learning algorithms is often a battle to balance out bias and variance.

* Increasing the bias will decrease the variance.
* Increasing the variance will decrease the bias.

In the next lesson you will discover the Linear Regression algorithm.

## Linear Regression Algorithm

Hi, linear regression is perhaps one of the most well known and well understood algorithms in statistics and machine learning.

Isn't it a technique from statistics?

Predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. We will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.

The representation of linear regression is a equation that describes a line that best fits the relationship between the input variables (x) and the output variables (y), by finding specific weightings for the input variables called coefficients (B).

For example:

y = B0 + B1 \* x

We will predict y given the input x and the goal of the linear regression learning algorithm is to find the values for the coefficients B0 and B1.

Different techniques can be used to learn the linear regression model from data, such as a linear algebra solution for ordinary least squares and gradient descent optimization.

Linear regression has been around for more than 200 years and has been extensively studied. Some good rules of thumb when using this technique are to remove variables that are very similar (correlated) and to remove noise from your data, if possible.

It is a fast and simple technique and good first algorithm to try.  
  
In the next lesson you will discover the Logistic Regression algorithm.  
  
Jason.  
  
Lesson 6: Logistic Regression Algorithm

Hi, logistic regression is another technique borrowed by machine learning from the field of statistics. It is the go-to method for binary classification problems (problems with two class values).

Logistic regression is like linear regression in that the goal is to find the values for the coefficients that weight each input variable.

Unlike linear regression, the prediction for the output is transformed using a non-linear function called the logistic function.

The logistic function looks like a big S and will transform any value into the range 0 to 1. This is useful because we can apply a rule to the output of the logistic function to snap values to 0 and 1 (e.g. IF less than 0.5 then output 1) and predict a class value.

Because of the way that the model is learned, the predictions made by logistic regression can also be used as the probability of a given data instance belonging to class 0 or class 1. This can be useful on problems where you need to give more rationale for a prediction.

Like linear regression, logistic regression does work better when you remove attributes that are unrelated to the output variable as well as attributes that are very similar (correlated) to each other.

It's a fast model to learn and effective on binary classification problems.  
  
In the next lesson you will discover the Linear Discriminant Analysis algorithm.

Lesson 7: Linear Discriminant Analysis Algorithm

Hi, Logistic Regression is a classification algorithm traditionally limited to only two-class classification problems. If you have more than two classes then the Linear Discriminant Analysis algorithm is the preferred linear classification technique.

The representation of LDA is pretty straight forward. It consists of statistical properties of your data, calculated for each class. For a single input variable this includes:

1. The mean value for each class.
2. The variance calculated across all classes.

Predictions are made by calculating a discriminate value for each class and making a prediction for the class with the largest value.

The technique assumes that the data has a Gaussian distribution (bell curve), so it is a good idea to remove outliers from your data before hand.

It's a simple and powerful method for classification predictive modeling problems.  
  
In the next lesson you will discover the Classification and Regression Trees algorithm.

# Lesson 8: Classification and Regression Trees

Hi, decision Trees are an important type of algorithm for predictive modeling machine learning.

The representation for the decision tree model is a binary tree. This is your binary tree from algorithms and data structures, nothing too fancy. Each node represents a single input variable (x) and a split point on that variable (assuming the variable is numeric).

The leaf nodes of the tree contain an output variable (y) which is used to make a prediction.  Predictions are made by walking the splits of the tree until arriving at a leaf node and output the class value at that leaf node.

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.

Decision trees have a high variance and can yield more accurate predictions when used in an ensemble, a topic we will cover in Lesson 13 and Lesson 14.  
  
In the next lesson you will discover the Naive Bayes algorithm