



Machine Learning Mastery

Making Developers Awesome at Machine Learning

[Click to Take the FREE Algorithms Crash-Course](#)



Linear Regression for Machine Learning

by **Jason Brownlee** on [March 25, 2016](#) in **Machine Learning Algorithms**

Tweet

Share

Share

Last Updated on August 12, 2019

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

In this post you will discover the linear regression algorithm, how it works and how you can best use it in on your machine learning projects. In this post you will learn:

- Why linear regression belongs to both statistics and machine learning.
- The many names by which linear regression is known.
- The representation and learning algorithms used to create a linear regression model.
- How to best prepare your data when modeling using linear regression.

You do not need to know any statistics or linear algebra to understand linear regression. This is a gentle high-level introduction to the technique to give you enough background to be able to use it effectively on your own problems.

Discover how machine learning algorithms work including kNN, decision trees, naive bayes, SVM, ensembles and much more [in my new book](#), with 22 tutorials and examples in excel.

Let's get started.



Linear Regression
Photo by Nicolas Raym

Isn't Linear Regression from

Before we dive into the details of linear regression, let's clarify what this algorithm.

Isn't it a technique from statistics?

Machine learning, more specifically the field of predictive modeling is primarily concerned with minimizing the error of a model or making the most accurate predictions possible, at the expense of explainability. In applied machine learning we will borrow, reuse and steal algorithms from many different fields, including statistics and use them towards these ends.

As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables, but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm.

Next, let's review some of the common names used to refer to a linear regression model.

Get your FREE Algorithms Mind Map

I've created a handy mind map of 60+ algorithms organized by type.

Download it, print it and use it.

Download

Your Start in Machine Learning

×

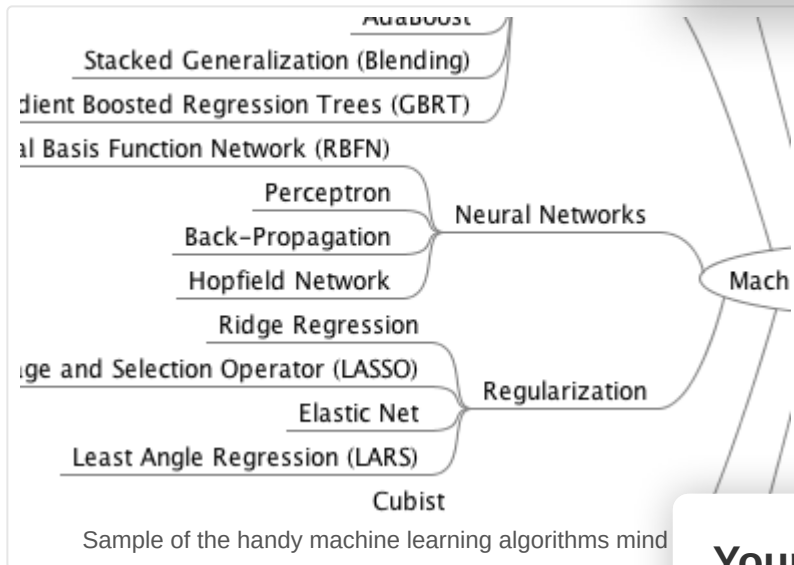
Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

Email Address

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE



Also get exclusive access to the machine learning algorithms email mini-course.

Many Names of Linear Regression

The reason is because linear regression has been studied from every possible angle and often

Linear regression is a **linear model**, e.g. a model with multiple input variables (x) and the single output variable (y). Most of the time, it's a combination of the input variables (x).

When there is a single input variable (x), the method is called **simple linear regression**. When there are **multiple input variables**, literature from the 1950s and 1960s called it **multiple regression**.

Different techniques can be used to prepare or train a linear regression model, the most common of which is called **Ordinary Least Squares**. It is common to therefore refer to a model prepared this way as **Ordinary Least Squares Linear Regression** or just **Least Squares Regression**.

Now that we know some names used to describe linear regression, let's take a closer look at the representation used.

Linear Regression Model Representation

[Linear regression](#) is an attractive model because the representation is so simple.

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.

The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

For example, in a simple regression problem (a single x and a single y), the form of the model would be:

Your Start in Machine Learning

$$y = B_0 + B_1x$$

In higher dimensions when we have more than one input (x), the line is called a plane or a hyper-plane. The representation therefore is the form of the equation and the specific values used for the coefficients (e.g. B_0 and B_1 in the above example).

It is common to talk about the complexity of a regression model like linear regression. This refers to the number of coefficients used in the model.

When a coefficient becomes zero, it effectively removes the influence of the input variable on the model and therefore from the prediction made from the model ($0 * x = 0$). This becomes relevant if you look at regularization methods that change the learning algorithm to reduce the complexity of regression models by putting pressure on the absolute size of

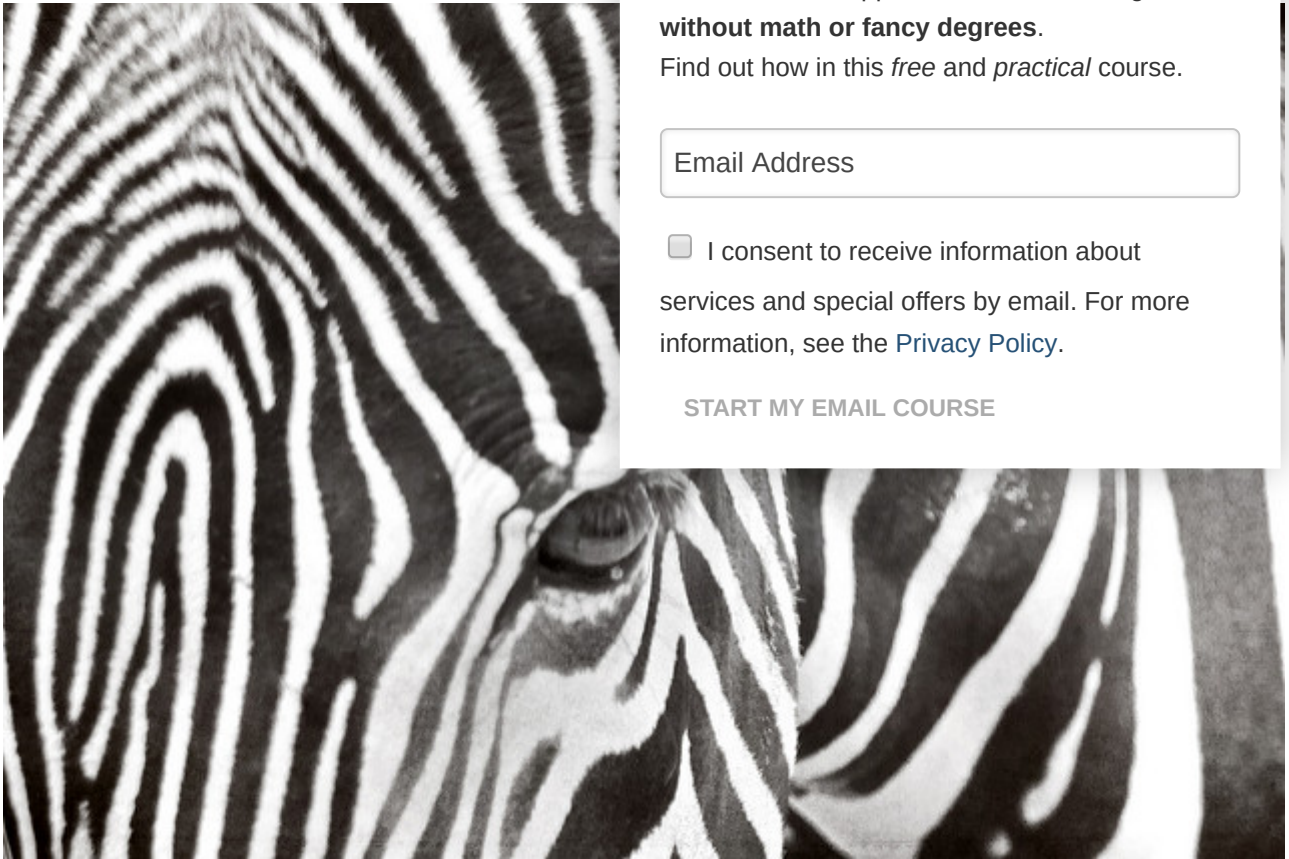
Now that we understand the representation used for the model, we can learn this representation from data.

Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE



What is Linear Regression?

Photo by [Estitxu Carton](#), some rights reserved.

Linear Regression Learning the Model

Learning a linear regression model means estimating the values of the coefficients used in the representation with the data that we have available.

In this section we will take a brief look at four techniques to prepare a linear regression model. This is not enough information to implement them from scratch, but enough to get a flavor of the computation and trade-offs involved.

Your Start in Machine Learning

There are many more techniques because the model is so well studied. Take note of Ordinary Least Squares because it is the most common method used in general. Also take note of Gradient Descent as it is the most common technique taught in machine learning classes.

1. Simple Linear Regression

With simple linear regression when we have a single input, we can use statistics to estimate the coefficients.

This requires that you calculate statistical properties from the data such as means, standard deviations, correlations and covariance. All of the data must be available to traverse and calculate statistics.

This is fun as an exercise in excel, but not really us

2. Ordinary Least Squares

When we have more than one input we can use O coefficients.

The [Ordinary Least Squares](#) procedure seeks to m means that given a regression line through the dat the regression line, square it, and sum all of the sq ordinary least squares seeks to minimize.

This approach treats the data as a matrix and uses values for the coefficients. It means that all of the c memory to fit the data and perform matrix operation

It is unusual to implement the Ordinary Least Squares procedure yourself unless as an exercise in linear algebra. It is more likely that you will call a procedure in a linear algebra library. This procedure is very fast to calculate.

3. Gradient Descent

When there are one or more inputs you can use a process of optimizing the values of the coefficients by iteratively minimizing the error of the model on your training data.

This operation is called [Gradient Descent](#) and works by starting with random values for each coefficient. The sum of the squared errors are calculated for each pair of input and output values. A learning rate is used as a scale factor and the coefficients are updated in the direction towards minimizing the error. The process is repeated until a minimum sum squared error is achieved or no further improvement is possible.

When using this method, you must select a learning rate (alpha) parameter that determines the size of the improvement step to take on each iteration of the procedure.

Gradient descent is often taught using a linear regression model because it is relatively straightforward to understand. In practice, it is useful when you have a very large dataset either in the number of rows or the number of columns that may not fit into mem

Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE

4. Regularization

There are extensions of the training of the linear model called regularization methods. These seek to both minimize the sum of the squared error of the model on the training data (using ordinary least squares) but also to reduce the complexity of the model (like the number or absolute size of the sum of all coefficients in the model).

Two popular examples of regularization procedures for linear regression are:

- **Lasso Regression:** where Ordinary Least Squares is modified to also minimize the absolute sum of the coefficients (called L1 regularization).
- **Ridge Regression:** where Ordinary Least Squares is modified to also minimize the squared absolute sum of the coefficients (called L2 regularization).

These methods are effective to use when there is a high degree of multicollinearity in the data, where ordinary least squares would overfit the training data.

Now that you know some techniques to learn the coefficients of the linear model, let's see how we can use a model to make predictions on new data.

Making Predictions with Linear Regression

Given the representation is a linear equation, making predictions is straightforward. We can plug in a specific set of inputs.

Let's make this concrete with an example. Imagine we have a linear regression model representation for this problem:

$$y = B_0 + B_1 * x_1$$

or

$$\text{weight} = B_0 + B_1 * \text{height}$$

Where B_0 is the bias coefficient and B_1 is the coefficient for the height column. We use a learning technique to find a good set of coefficient values. Once found, we can plug in different height values to predict the weight.

For example, let's use $B_0 = 0.1$ and $B_1 = 0.5$. Let's plug them in and calculate the weight (in kilograms) for a person with the height of 182 centimeters.

$$\text{weight} = 0.1 + 0.5 * 182$$

$$\text{weight} = 91.1$$

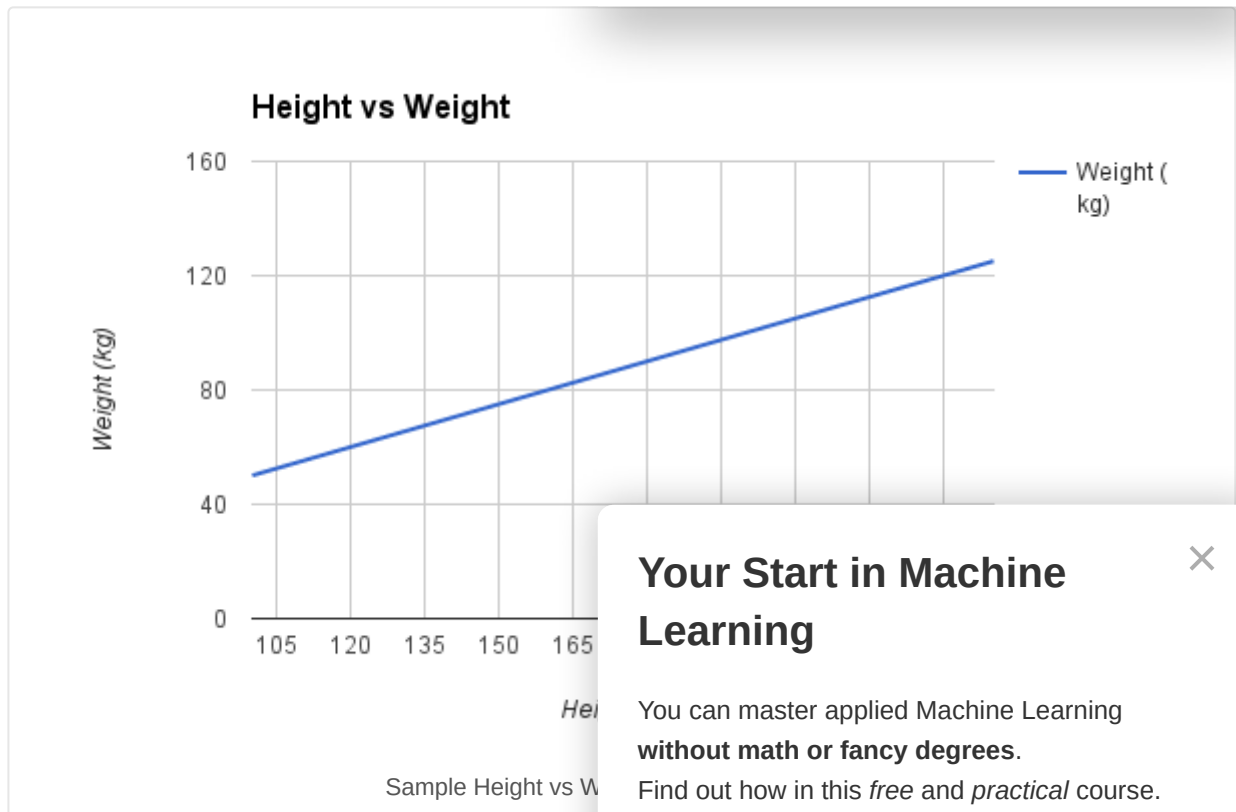
You can see that the above equation could be plotted as a line in two-dimensions. The B_0 is our starting point regardless of what height we have. We can run through a bunch of heights from 100 to 250 centimeters and plug them to the equation and get weight values, creating our line.

Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE



Now that we know how to make predictions given a set of rules of thumb for preparing our data to make the model

Preparing Data For Linear Regression

Linear regression is been studied at great length, and data should be structured to make best use of the model.

As such, there is a lot of sophistication when talking about these requirements and expectations which can be intimidating. In practice, you can use these rules more as rules of thumb when using Ordinary Least Squares Regression, the most common implementation of linear regression.

Try different preparations of your data using these heuristics and see what works best for your problem.

- **Linear Assumption.** Linear regression assumes that the relationship between your input and output is linear. It does not support anything else. This may be obvious, but it is good to remember when you have a lot of attributes. You may need to transform data to make the relationship linear (e.g. log transform for an exponential relationship).
- **Remove Noise.** Linear regression assumes that your input and output variables are not noisy. Consider using data cleaning operations that let you better expose and clarify the signal in your data. This is most important for the output variable and you want to remove outliers in the output variable (y) if possible.
- **Remove Collinearity.** Linear regression will over-fit your data when you have highly correlated input variables. Consider calculating pairwise correlations for your input data and removing the most correlated.
- **Gaussian Distributions.** Linear regression will make more reliable predictions if your input and output variables have a Gaussian distribution. You may get some benefit using transforms (e.g. log or BoxCox) on your variables to make their distributions more Gaussian.

Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE

Your Start in Machine Learning

- **Rescale Inputs:** Linear regression will often make more reliable predictions if you rescale input variables using standardization or normalization.

See the [Wikipedia article on Linear Regression](#) for an excellent list of the assumptions made by the model. There's also a great list of assumptions on the [Ordinary Least Squares Wikipedia article](#).

Further Reading

There's plenty more out there to read on linear regression. Start using it before you do more reading, but when you want to dive deeper, below are some references you could use.

Machine Learning Books that Mention Linear Regression

These are some machine learning books that you can find [here](#). I mention linear regression in the context of machine learning.

- [A First Course in Machine Learning](#), Chapter 10.
- [An Introduction to Statistical Learning: with Applications in R](#), Chapter 10.
- [Applied Predictive Modeling](#), Chapter 6.
- [Machine Learning in Action](#), Chapter 8.
- [The Elements of Statistical Learning: Data Mining, Inference, and Prediction](#), Chapter 3.

Posts on Linear Regression

Below are some interesting essays and blog posts on linear regression.

- [Ordinary Least Squares Regression: Explained](#)
- [Ordinary Least Squares Linear Regression: Flaws, Problems and Pitfalls](#)
- [Introduction to linear regression analysis](#)
- [Four Assumptions Of Multiple Regression That Researchers Should Always Test](#)

Know any more good references on linear regression with a bent towards machine learning and predictive modeling? Leave a comment and let me know.

Summary

In this post you discovered the linear regression algorithm for machine learning.

You covered a lot of ground including:

- The common names used when describing linear regression models.
- The representation used by the model.
- Learning algorithms used to estimate the coefficients in the model.
- Rules of thumb to consider when preparing data for use with linear regression.

Try out linear regression and get comfortable with it.

Do you have any questions about linear regression or about this post?

Leave a comment and ask, I will do my best to answer.

Your Start in Machine Learning

You can master applied Machine Learning **without math or fancy degrees**. Find out how in this *free* and *practical* course.

☐ I consent to receive information about services and special offers by email. For more information, see the [Privacy Policy](#).

START MY EMAIL COURSE