**Machine Learning:**

**Linear Regression in one variable:**

There are two ways of expressing the output 1) as a continuous real valued number or 2) by classifying it into any one of the ‘n’ finite classes. The former method is called Regression and the latter one is called Classification.

**Model function:**

Let X be a set of input features and Y be its corresponding output that we are trying to predict.

A model function (h) is a function which on the training set that tries to provide a prediction for every input feature X; i.e.

h: X -> Y

An example for a linear model function:

hXiXi

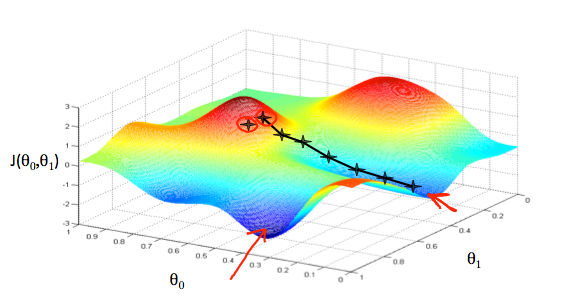
**Cost function:**

It’s a function defined to calculate the error or the prediction accuracy of the defined model function. Its defined as:

J() = (h yi)2/2m(m is the number of training data variables)

The aim of the learning algorithm is thus to reduce this function as much as possible.

When the above function is plotted, w.r.t to its two parameters (and ) and the output J() a 3-dimensional plot like the one below is obtained.



The aim is to locate the values (marked with red arrows), one method is called the gradient descent method.

**Gradient Descent Method:**

In gradient descent method an arbitrary point is chosen, then a small step is taken in the direction that has the greatest downward slope. After reaching the next point, the same process is repeated until a valley is reached. A property known with the gradient descent method is that it finds the nearest (local) minimum w.r.t to the starting point.

Formula for gradient descent:

J

J

The above process is repeated until convergence (the output becomes same after the next iteration).

Substituting equation for the cost function into the gradient descent equation:

mhXi-Yisummation from 1 to m

mhXi-YiXisummation from 1 to m