### Learning Algorithm

Linear regression is trained by minimizing a cost function over the entire training set.

The cost function J for a particular choice of parameters Ѳ is the mean squared error (MSE):

Where the variables used are

*m* The number of training examples

*x(i)* The input vector for the ith training example

*y(i)*The class label for the ith training example

Ѳ The chosen parameter values or “weights” (Ѳ0, Ѳ1, Ѳ2)

The algorithm’s prediction for the ith training example using the parameters Ѳ.

The cost function can be thought of a measure of the classifier’s performance on the training set. The cost is higher when the classifier is performing poorly on the training set. The objective of the learning algorithm, then, is to find the parameters Ѳ which give the minimum possible cost J.

For this paper, we’ve used batch gradient descent to iteratively adjust the parameters Ѳ. Each iteration of gradient descent makes a small adjustment to the parameters Ѳ which reduces the cost (the MSE).

Because the cost is averaged over the dataset, the outlying data points will have a smaller influence over the final weights than the dense points.