Cost Function - Intuition I

We want to fit a straight line to our data.

Hypothesis: $h_\theta(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$

Goal: $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

In order to better visualize the cost function $J(\theta_0, \theta_1)$, we're going to work with a simplified hypothesis function.

Simplified (Assumes $\theta_0 = 0$):

Hypothesis: $h_\theta(x) = \theta_1 x$

Parameters: θ_1

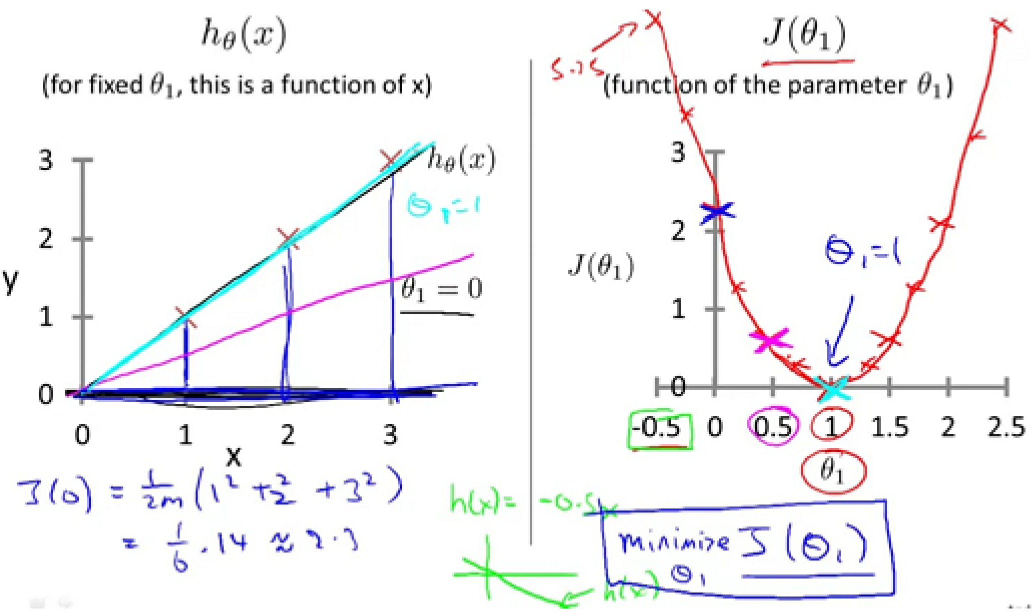
Cost Function: $J(\theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$

Goal: $\min_{\theta_0} J(\theta_0)$

$h_\theta(x)$ for fixed θ_1, this is a function of x

$J(\theta_1)$ function of the parameter θ_1

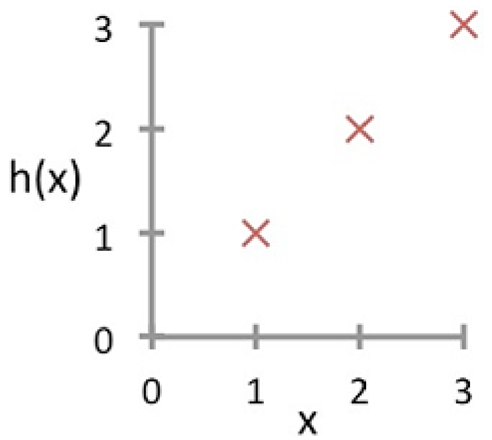
Each value of theta one corresponds to a different hypothesis, or to a different straight line fit on the left. And for each value of theta one, we could then derive a different value of j of theta one.



- **Hypothesis** - Is like your prediction machine, throw in an x value, get a putative y value.
- **Cost** - Is a way to, using your training data, determine values for your θ values which make the hypothesis as accurate as possible.

The optimization objective for the learning algorithm is find the value of θ_1 which minimizes $J(\theta_1)$

Video Question: Suppose we have a training set with $m = 3$ examples, plotted below. Our hypothesis representation is $h_\theta(x) = \theta_1 x$, with parameter θ_1. The cost function $J(\theta_1)$ is $J(\theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$. What is $J(0)$?



- 0
- 1/6
- 1
- 14/6

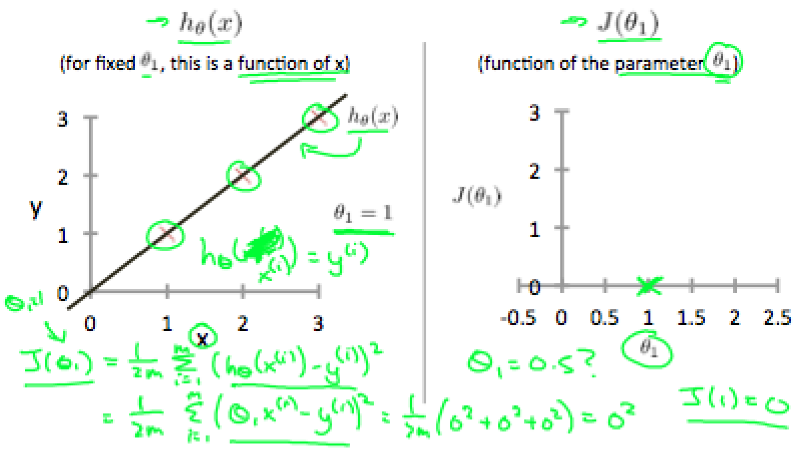
Solution: $J(0) = \frac{1}{2m}[(0-1)^2 + (0-2)^2 + (0-3)^2] = \frac{1}{2(3)}[14] = \frac{14}{6}$

Summary

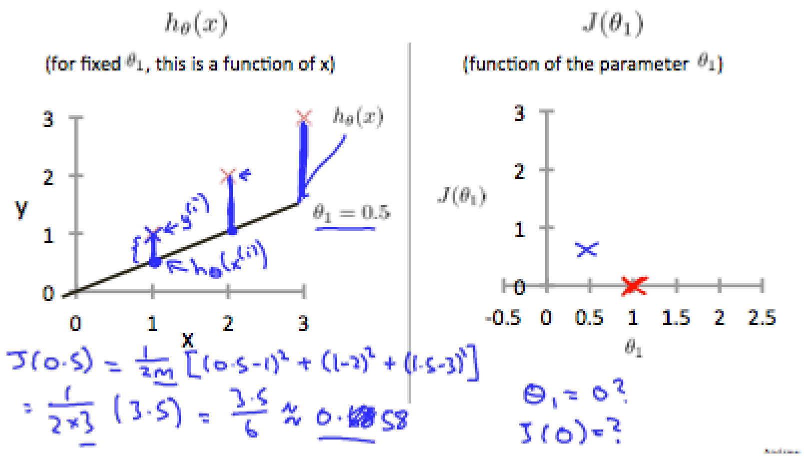
If we try to think of it in visual terms, our training data set is scattered on the x-y plane. We are trying to make a straight line (defined by $h_\theta(x)$) which passes through these scattered data points.

Our objective is to get the best possible line. The best possible line will be such so that the average squared vertical distances of the scattered points from the line will be the least.

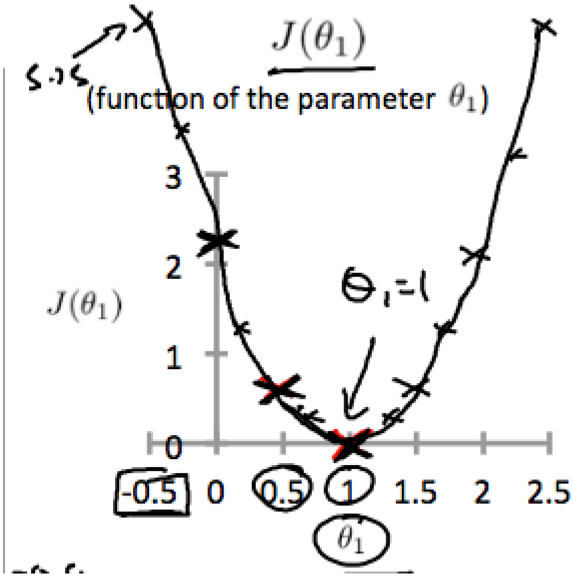
Ideally, the line should pass through all the points of our training data set. In such a case, the value of $J(\theta_0, \theta_1)$ will be 0. The following example shows the ideal situation where we have a cost function of 0.



When $\theta_1 = 1$, we get a slope of 1 which goes through every single data point in our model. Conversely, when $\theta_1 = 0.5$, we see the vertical distance from our fit to the data points increase.



This increases our cost function to 0.58. Plotting several other points yields to the following graph:

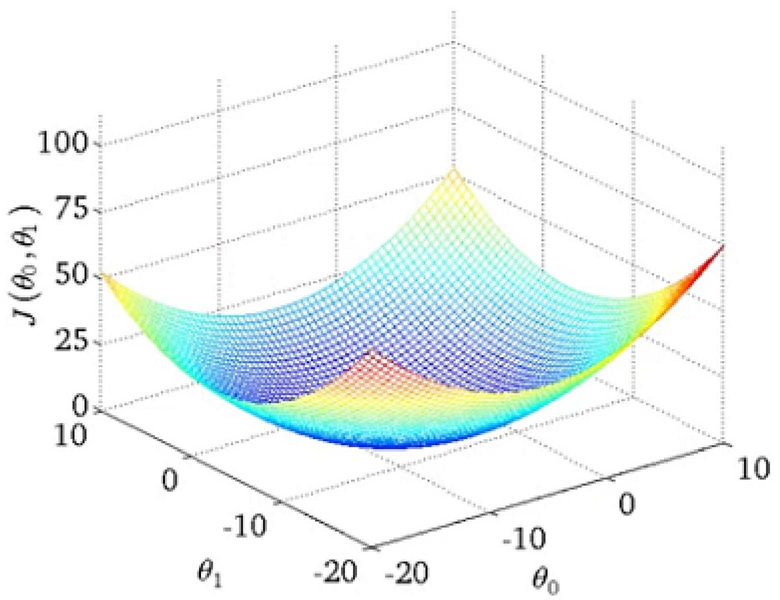


Thus as a goal, we should try to minimize the cost function. In this case, $\theta_1 = 1$ is our global minimum.

Cost Function - Intuition II

When we have only one parameter, the parts we drew had the sort of bow shaped function. Now, when we have two parameters, it turns out the cost function has a sort of bow shape. The graph below is a 3-D surface plot, where the axes are labeled theta zero and theta one.

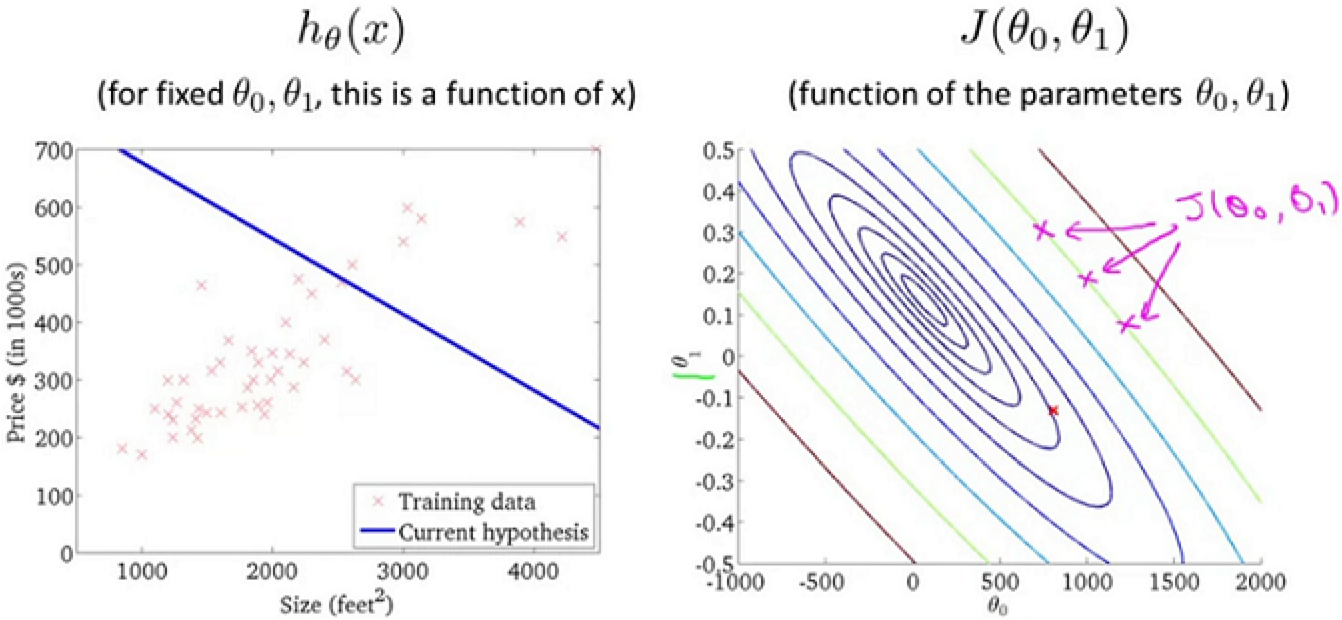
So as we vary theta zero and theta one, the two parameters, we get different values of the cost function J (theta zero, theta one) and the height of the surface above a particular point of theta zero, theta one.



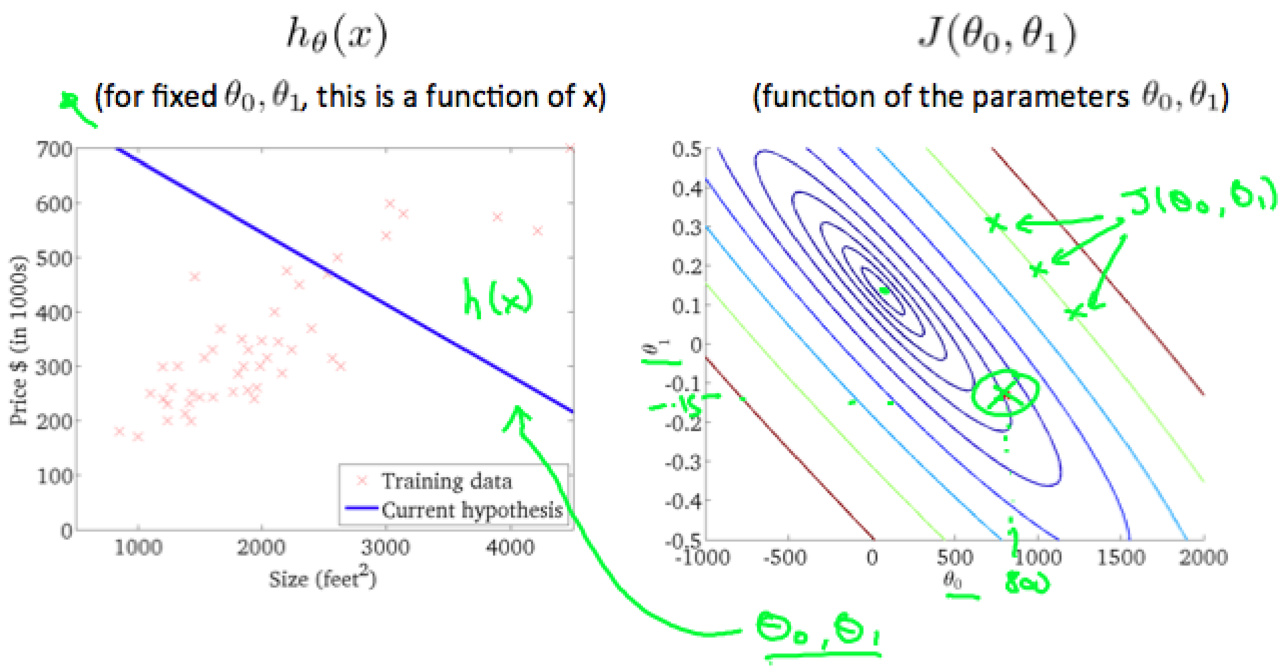
Contour plots (contour figures)

The contour figures is a more convenient way to visualize the function J. As we vary theta zero and theta one, the two parameters, we get different values of the cost function $J(\theta_0, \theta_1)$ and the height of the surface above a particular point of theta zero, theta one.

What each of the ovals shows, what each of the ellipsis shows is a set of points that takes on the same value for $J(\theta_0, \theta_1)$.

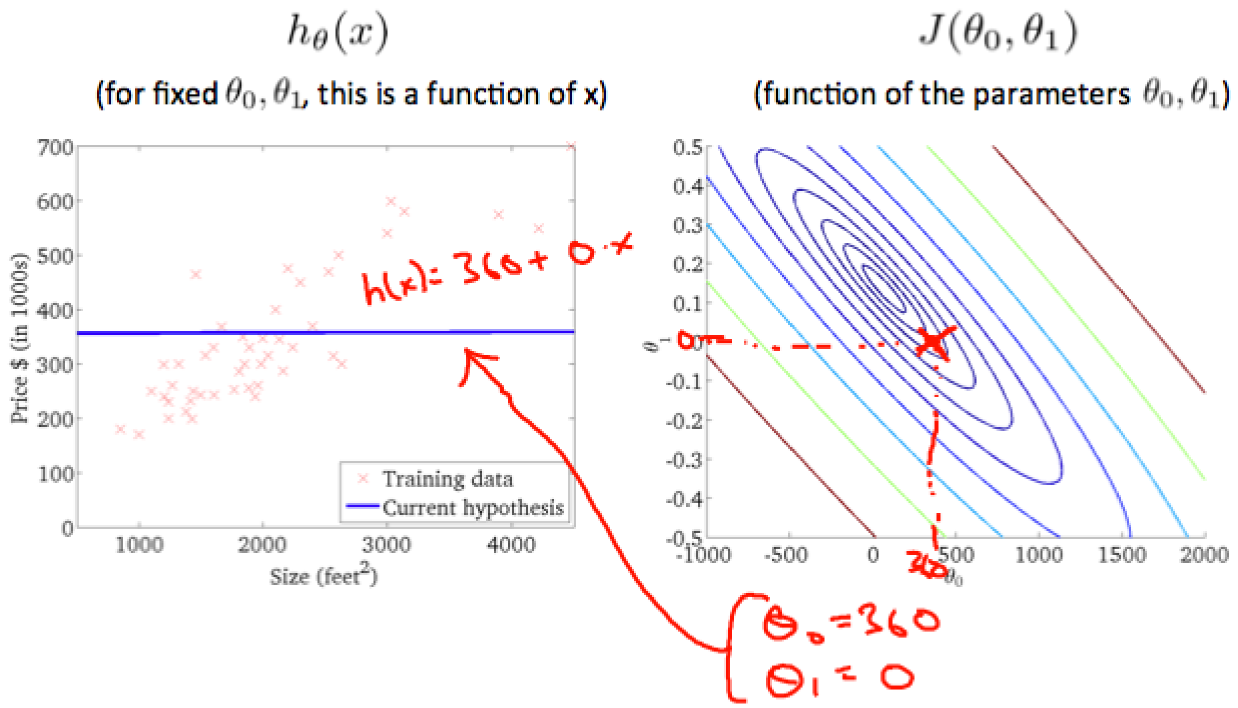


A contour plot is a graph that contains many contour lines. A contour line of a two variable function has a constant value at all points of the same line. An example of such a graph is the one to the right below.

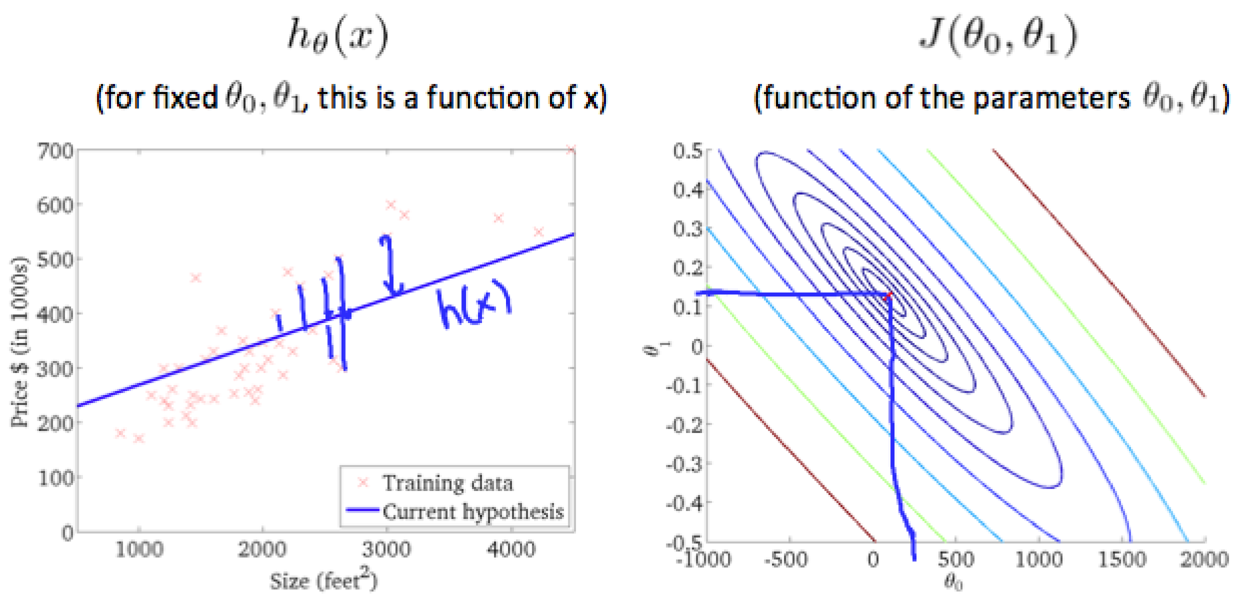


Taking any color and going along the 'circle', one would expect to get the same value of the cost function. For example, the three green points found on the green line above have the same value for $J(\theta_0, \theta_1)$ and as a result, they are found along the same line.

The circled x displays the value of the cost function for the graph on the left when $\theta_0 = 800$ and $\theta_1 = -0.15$. Taking another $h(x)$ and plotting its contour plot, one gets the following graphs:



When $\theta_0 = 360$ and $\theta_1 = 0$, the value of $J(\theta_0, \theta_1)$ in the contour plot gets closer to the center thus reducing the cost function error. Now giving our hypothesis function a slightly positive slope results in a better fit of the data.



The graph above minimizes the cost function as much as possible and consequently, the result of θ_0 and θ_1 tend to be around 0.12 and 250 respectively. Plotting those values on our graph to the right seems to put our point in the center of the inner most 'circle'.