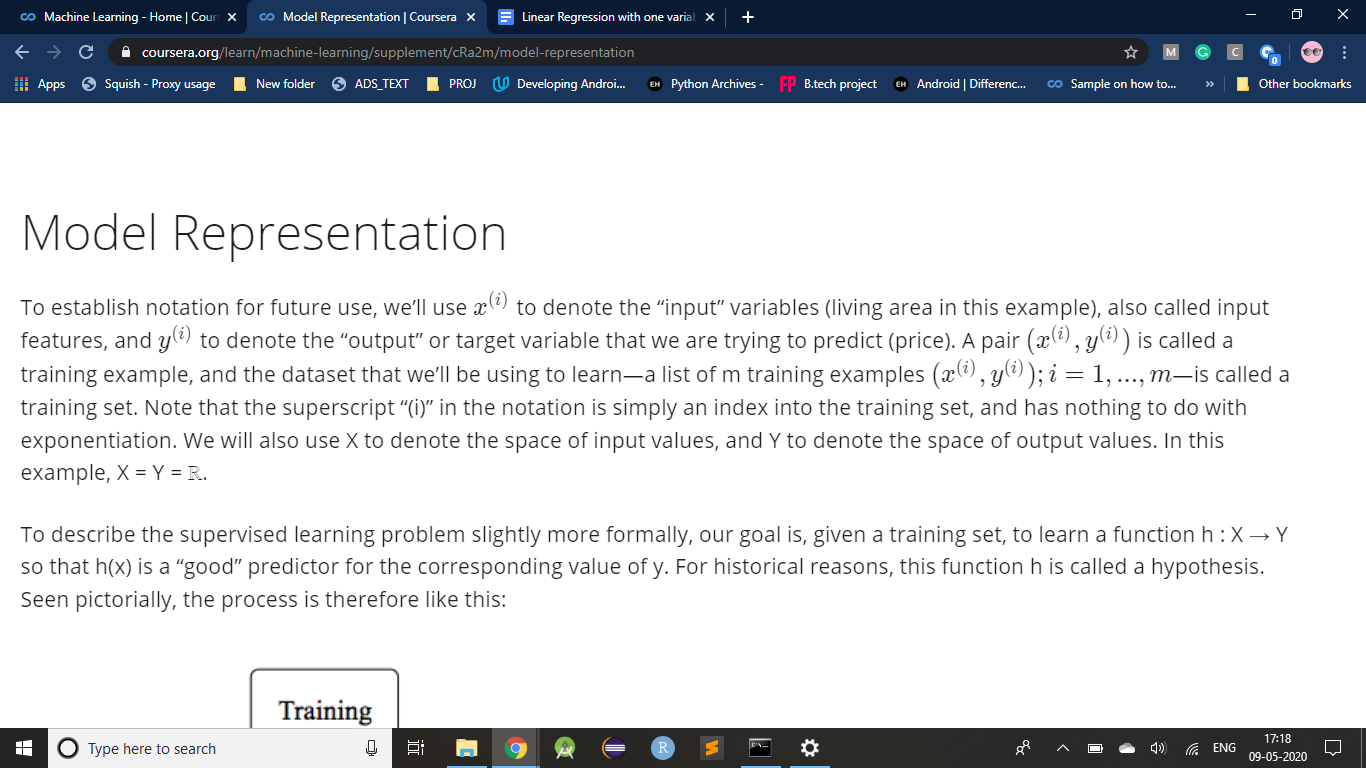
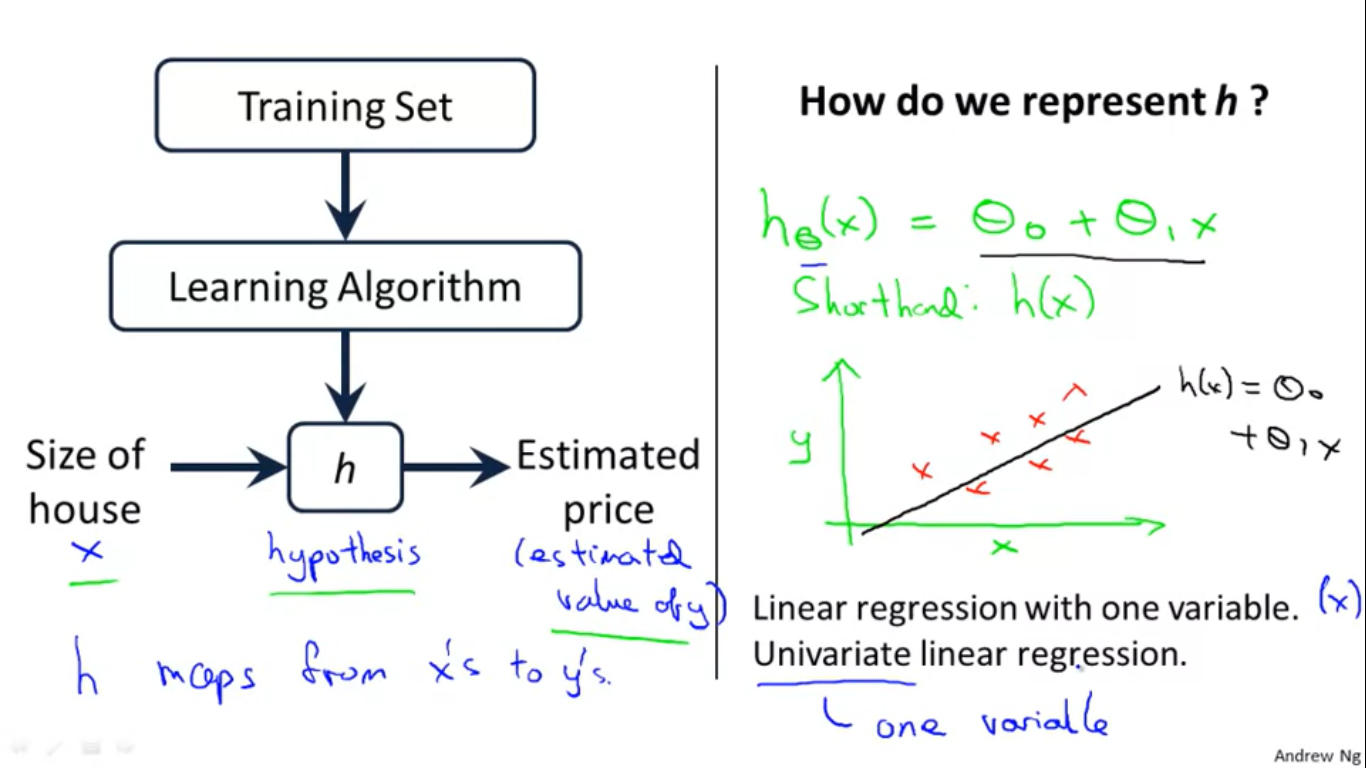
Model Representation

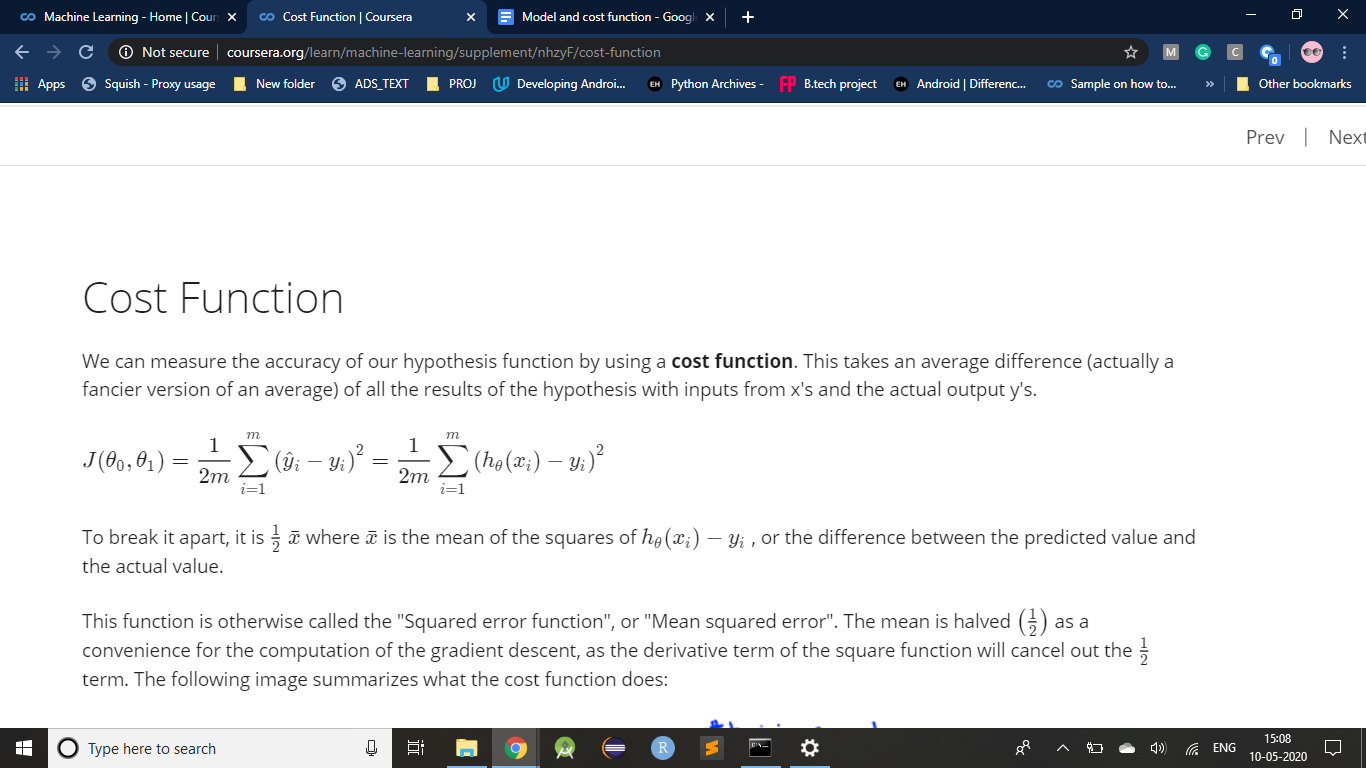


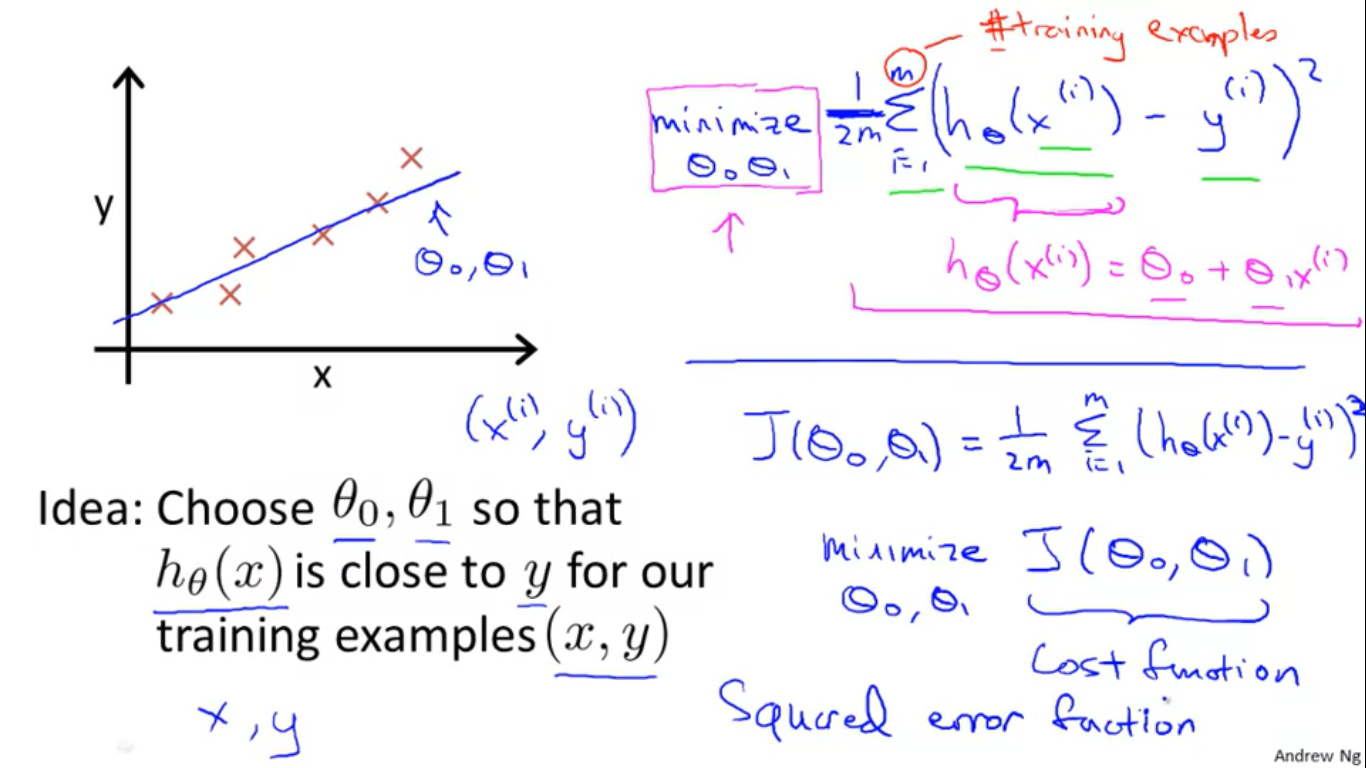


When the target variable that we’re trying to predict is continuous, such as in our housing example, we call the learning problem a regression problem.

When y can take on only a small number of discrete values (such as if, given the living area, we wanted to predict if a dwelling is a house or an apartment, say), we call it a classification problem.

Cost Function

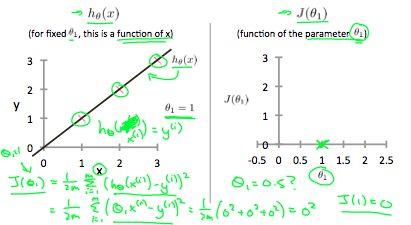




Cost Function(Intuition 1) :

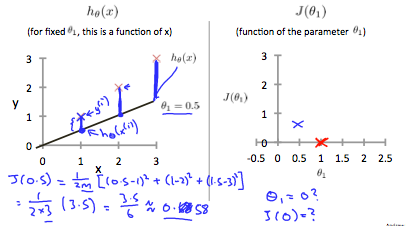
If we try to think of it in visual terms, our training data set is scattered on the x-y plane. We are trying to make a straight line (defined by h θ(x)) which passes through these scattered data points.

Our objective is to get the best possible line. The best possible line will be such 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(θ 0,θ 1) will be 0. The following example shows the ideal situation where we have a cost function of 0.

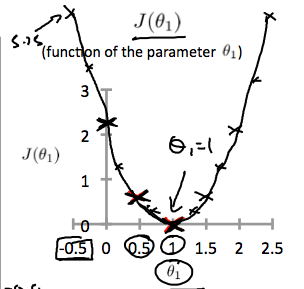


When *θ*1 = 1, we get a slope of 1 which goes through every single data point in our model. Conversely, when

*θ*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 :

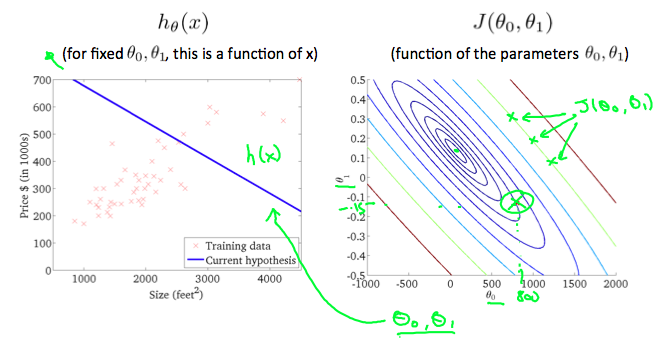


Thus as a goal, we should try to minimize the cost function. In this case, *θ*1=1 is our global minimum.

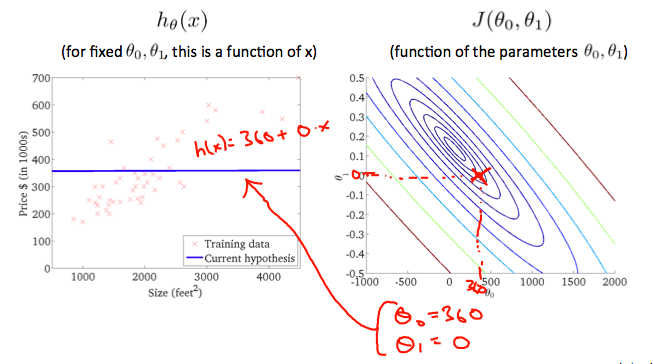
Cost Function(Intuition 2) :

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. They are a way of showing 3D figures on 2D planes.

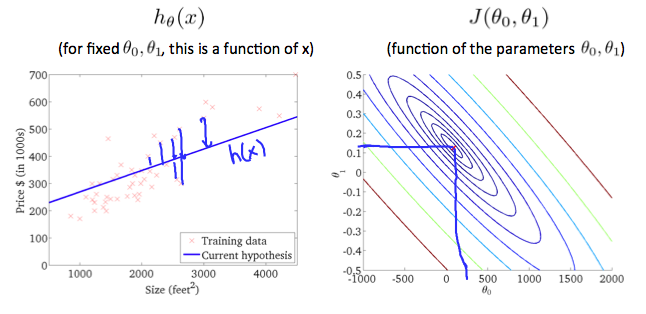
An example of such a graph is the one to the right below.



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*(*θ*0,*θ1*) 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 *θ0* = 800 and *θ*1= -0.15. Taking another h(x) and plotting its contour plot, one gets the following graphs:



When *θ*0= 360 and *θ*1= 0, the value of *J*(*θ*0,*θ*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 *θ1* and *θ*0 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 innermost 'circle'.