## Cost Function (Squared Error Function)

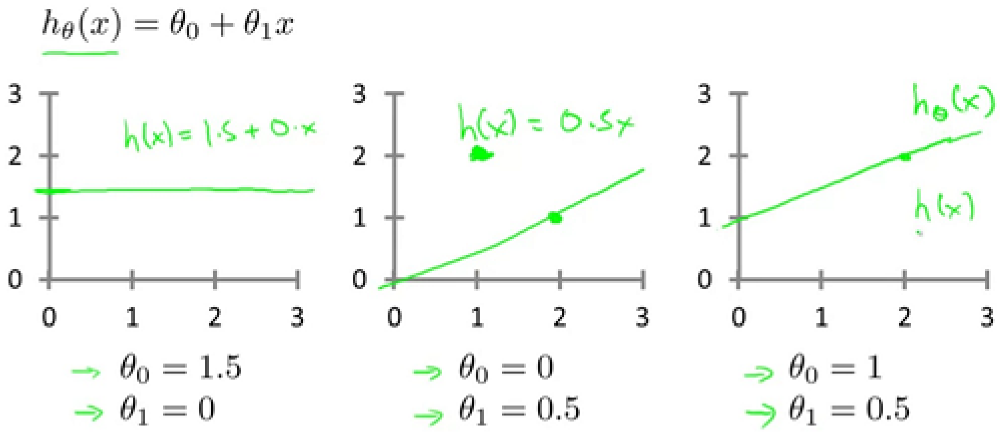
In linear progression, we have a training set. And the form of our hypothesis ($h_\theta$), which we use to make predictions is the linear function: $h_\theta(x) = \theta_0 + \theta_1 x$.

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

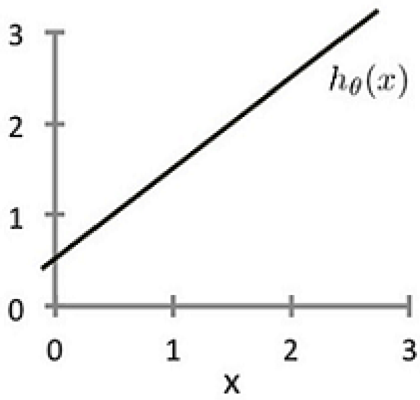
$\theta_i$'s: Parameters

How to choose $\theta_i$'s ?

With different choices of the parameter's theta 0 and theta 1, we get different hypothesis, different hypothesis functions.




**Video Question:** Consider the plot below of $h_\theta(x) = \theta_0 + \theta_1 x$. What are $\theta_0$ and $\theta_1$?




$\theta_0 = 0, \theta_1 = 1$

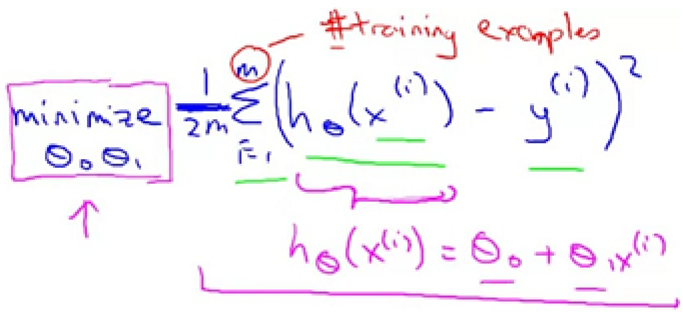
$\theta_0 = 0.5, \theta_1 = 1$

$\theta_0 = 1, \theta_1 = 0.5$

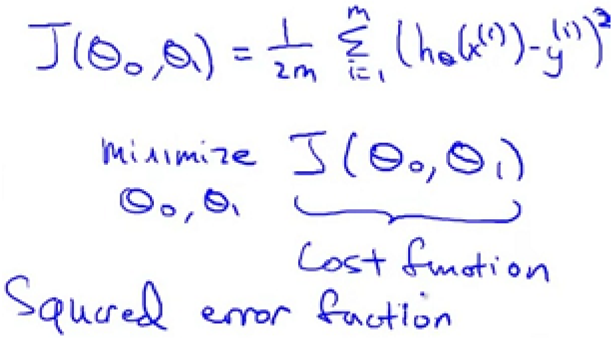
$\theta_0 = 1, \theta_1 = 1$

In linear regression, we have a training set and what we want to do, is come up with values for the parameters $\theta_0$ and $\theta_1$ so that the straight line we get, corresponds to a straight line that somehow fits the data well. So, how do we come up with values, $\theta_0, \theta_1$, that corresponds to a good fit to the data? The idea is we get to choose our parameters $\theta_0, \theta_1$ so that $h_\theta(x)$ is close to "y" for our training examples $(x, y)$. In our training set, we've given a number of examples where we know X decides the wholes and we know the actual price is was sold for.

In linear regression, what we're going to do is, we're going to want to solve a **minimization problem**. We want to minimize the sum of my training set, sum from i = 1, ..., M, of the difference of the squared error, the square difference between the predicted price of a house, and the price that it was actually sold for.




By convention we usually define a **cost function**, and what we want to do is minimize over $\theta_0$ and $\theta_1$. This cost function is also called the **squared error function**. When sometimes called the **squared error cost function** and it turns out that why do we take the squares of the erros. It turns out that the squared error cost function is a reasonable choice and works well for problems for most regression programs.




## Summary

We can measure the accuracy of our hypothesis function by using a cost function. This takes an average difference (actually a fancier version of an average) of all the results of the hypothesis with inputs from x's and the actual output y's.

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2 = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta(x_i) - y_i\right)^2$$

To break it apart, it is $\frac{1}{2}\bar{x}$ where $\bar{x}$ is the mean of the squares of $h_\theta(x_i) - y_i$, or the difference between the predicted value and the actual value. This function is otherwise called the **"Squared error function"**, or **"Mean squared error"**. The mean is halved $\left(\frac{1}{2}\right)$ as a convenience for the computation of the **gradient descent**, as the derivative term of the square function will cancel out the $\frac{1}{2}$ term. The following image summarizes what the cost function does: