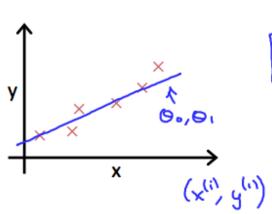
Cost Function

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.

$$J(heta_0, heta_1) = rac{1}{2m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 = rac{1}{2m} \sum_{i=1}^m \left(h_{ heta}(x_i) - y_i
ight)^2$$

To break it apart, it is $rac{1}{2}$ $ar{x}$ where $ar{x}$ is the mean of the squares of $h_{ heta}(x_i)-y_i$, 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 $rac{1}{2}$ term. The following image summarizes what the cost function does:



Idea: Choose θ_0, θ_1 so that $h_{\theta}(x)$ is close to y for our training examples (x, y)

× ,4

Miximize J(00,01)
00,01 Cost function
Squared error faction

✓ Complete

