

Welcome

Introduction

- ✓ **Video:** Welcome
6 min
- ✓ **Video:** What is Machine Learning?
7 min
- ✓ **Reading:** What is Machine Learning?
5 min
- ✓ **Reading:** How to Use Discussion Forums
4 min
- ✓ **Video:** Supervised Learning
12 min
- ✓ **Reading:** Supervised Learning
4 min
- ✓ **Video:** Unsupervised Learning
14 min
- ✓ **Reading:** Unsupervised Learning
3 min
- ✓ **Reading:** Who are Mentors?
3 min
- ✓ **Reading:** Get to Know Your Classmates
8 min
- ✓ **Reading:** Frequently Asked Questions
11 min

Review

- ✓ **Reading:** Lecture Slides
20 min

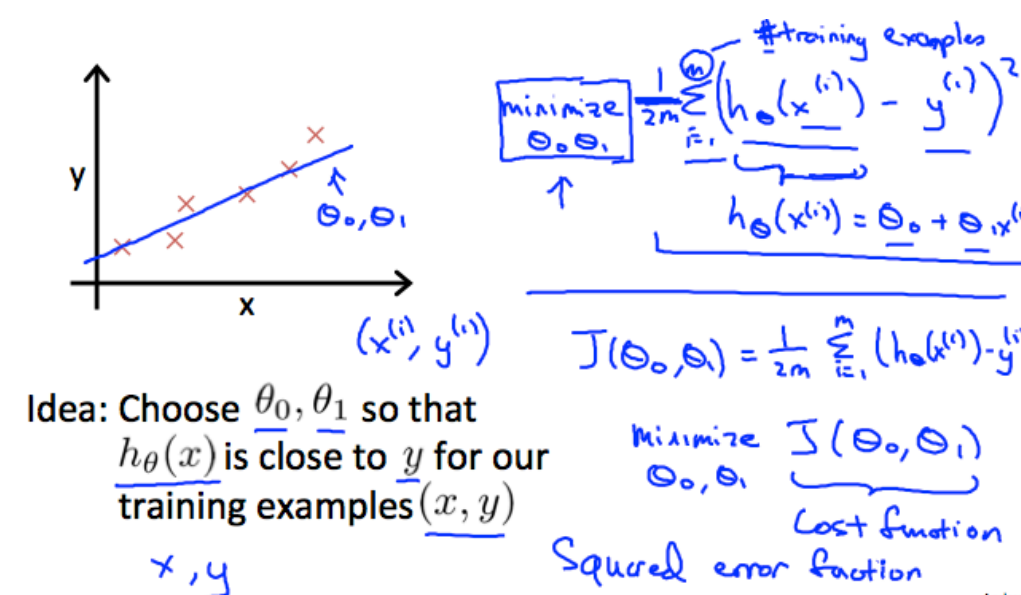
Cost Function

We can measure the accuracy of our hypothesis function h_θ 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.

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

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 ($\frac{1}{2}$) 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:



✓ Complete

Go to next item