

Welcome

- ✓ **Video:** Welcome to Machine Learning! 1 min
- ✓ **Reading:** Machine Learning Honor Code 8 min

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
- ✓ **Quiz:** Introduction 5 questions

Model and Cost Function

- ✓ **Video:** Model Representation 8 min
- ✓ **Reading:** Model Representation

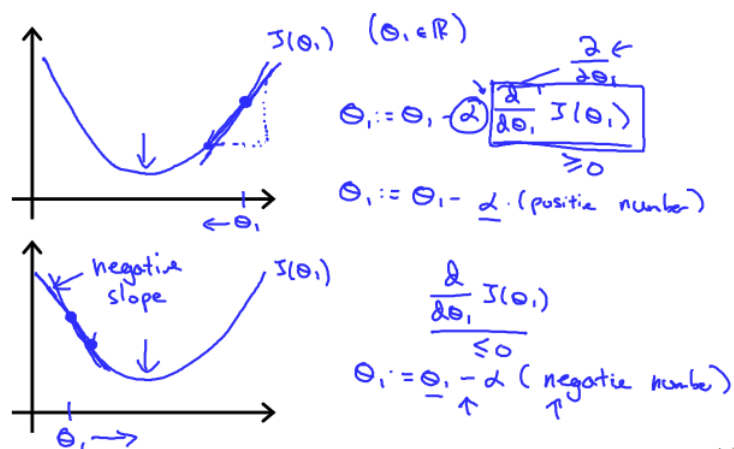
Gradient Descent Intuition

In this video we explored the scenario where we used one parameter θ_1 and plotted its cost function to implement a gradient descent. Our formula for a single parameter was :

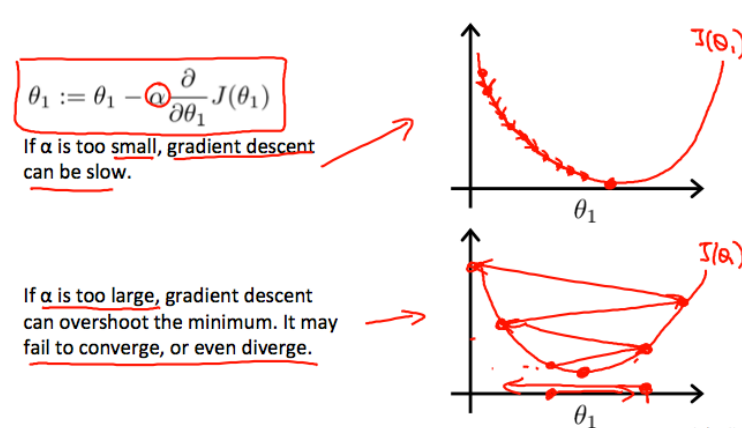
Repeat until convergence:

$$\theta_1 := \theta_1 - \alpha \frac{d}{d\theta_1} J(\theta_1)$$

Regardless of the slope's sign for $\frac{d}{d\theta_1} J(\theta_1)$, θ_1 eventually converges to its minimum value. The following graph shows that when the slope is negative, the value of θ_1 increases and when it is positive, the value of θ_1 decreases.



On a side note, we should adjust our parameter α to ensure that the gradient descent algorithm converges in a reasonable time. Failure to converge or too much time to obtain the minimum value imply that our step size is wrong.



How does gradient descent converge with a fixed step size α ?

The intuition behind the convergence is that $\frac{d}{d\theta_1} J(\theta_1)$ approaches 0 as we approach the bottom of our convex function. At the minimum, the derivative will always be 0 and thus we get:

$$\theta_1 := \theta_1 - \alpha * 0$$