

Welcome

Introduction

Review

Model and Cost Function

✓ **Video:** Model Representation
8 min

✓ **Reading:** Model Representation
3 min

✓ **Video:** Cost Function
8 min

✓ **Reading:** Cost Function
3 min

▶ **Video:** Cost Function - Intuition I
11 min

📖 **Reading:** Cost Function - Intuition I
4 min

▶ **Video:** Cost Function - Intuition II
8 min

📖 **Reading:** Cost Function - Intuition II
3 min

Parameter Learning

▶ **Video:** Gradient Descent
11 min

📖 **Reading:** Gradient Descent
3 min

▶ **Video:** Gradient Descent Intuition
11 min

📖 **Reading:** Gradient Descent Intuition
3 min

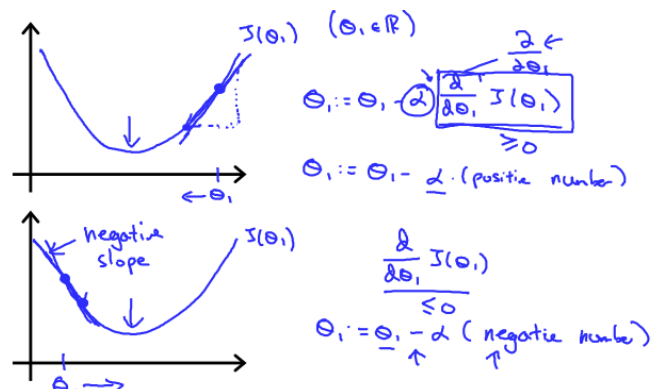
Gradient Descent Intuition

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.

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

If α is too small, gradient descent can be slow.

If α is too large, gradient descent can overshoot the minimum. It may fail to converge, or even diverge.

