Audit Access Expires May 11, 2020
You lose all access to this course, including your progress, on May 11, 2020. Upgrade by Mar 25, 2020 to get unlimited access to the course as long as it exists on the site. **Upgrade now**

4. Gradient Descent
Gradient Descent

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Gradient Descent: Geometrically Revisited

2/2 points (graded)
Assume $\theta \in \mathbb{R}$. Our goal is to find θ that minimizes

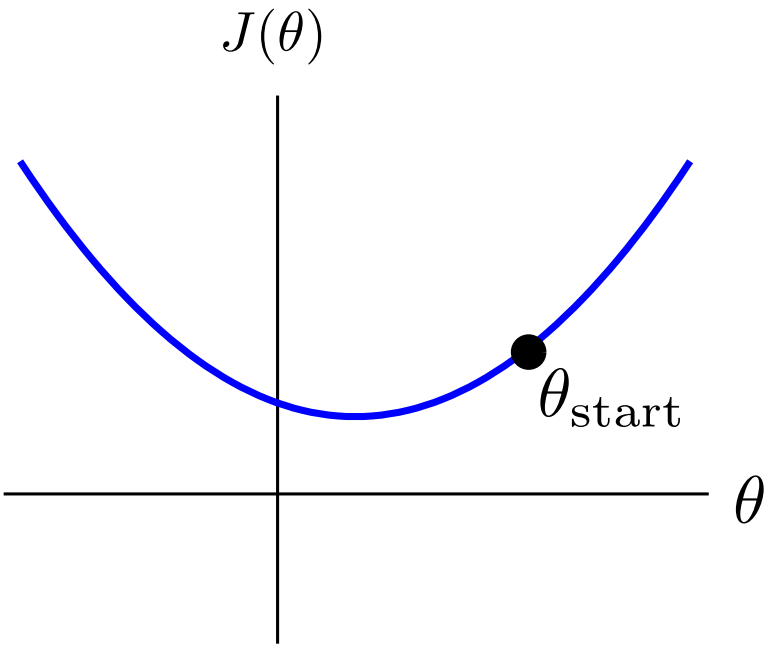
$$J\left(\theta, \theta_0\right) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h \left(y^{(i)} \left(\theta \cdot x^{(i)} + \theta_0\right)\right) + \frac{\lambda}{2} \mid\mid \theta \mid\mid^2$$

through gradient descent. In other words, we will

1. Start θ at an arbitrary location: $\theta \leftarrow \theta_{start}$
2. Update θ repeatedly with $\theta \leftarrow \theta - \eta \frac{\partial J(\theta, \theta_0)}{\partial \theta}$ until θ does not change significantly

In the 2 dimensional space below, we start our gradient descent at θ_{start}. What is

the direction θ moves to in its first update?



- away from the origin
- towards the origin
- upwards
- downwards

✔

What happens if we increase the stepsize η?

- the magnitude of change in each update gets larger
- the magnitude of change in each update gets smaller

Solution:

Gradient descent makes θ move to opposite direction of the gradient. Thus it will move towards the origin at θ_{start}. Also, increasing the stepsize makes the update happen in greater magnitude.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Lecture 4. Linear Classification and Generalization / 4. Gradient Descent

Add a Post

Show all posts by recent activity

- Theta_0 not included. Why? 3
- The graph in the question 3
- Parameter theta_0 2
  The parameter theta_0 is not taken into account in the task.

Learn About Verified Certificates

© All Rights Reserved