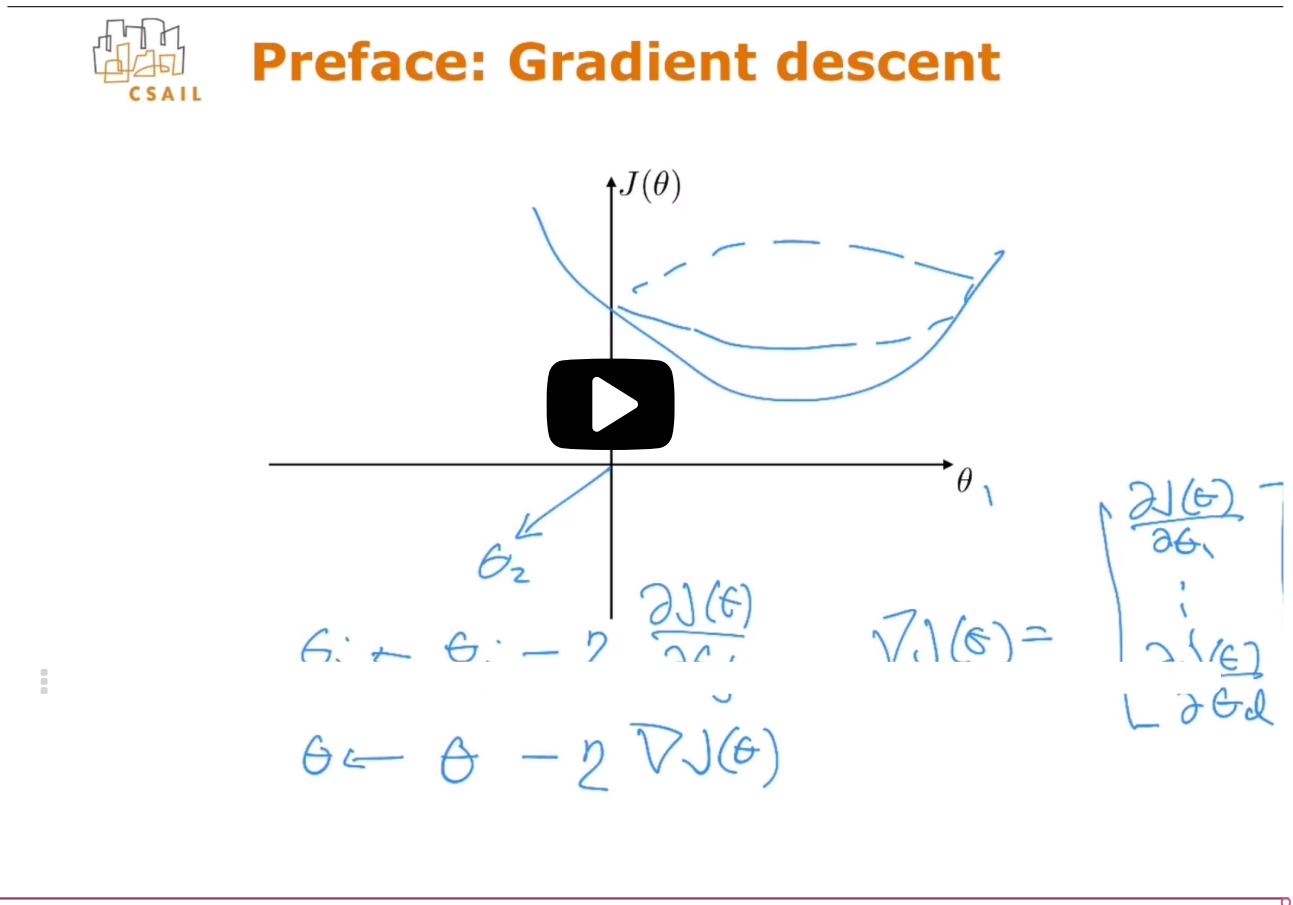
4. Gradient Descent
Gradient Descent



4:19 / 4:19 1.25x

the parameters.

And this gradient is nothing but a concatenation

of these individual derivatives with respect to the parameters.

So derivative of the function with respect to the first

coordinate and then derivative of the function with respect

to the last coordinate-- call it theta d--

of the d coordinates in theta.

End of transcript. Skip to the start.

Video
Download video file

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

Gradient Descent: Geometrically Revisited

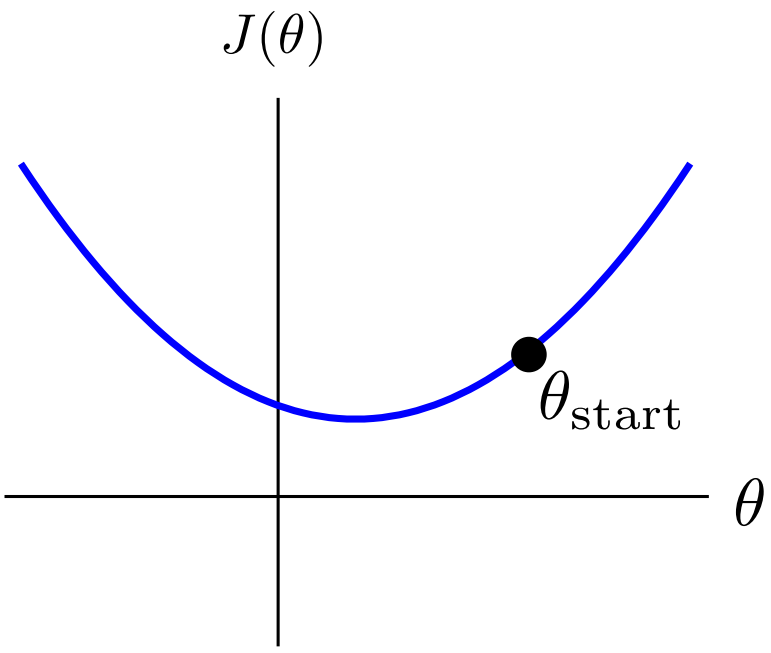
2/2 points (graded)
Assume $\theta \in \mathbb{R}$. Our goal is to find θ that minimizes

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

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

Show Discussion

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