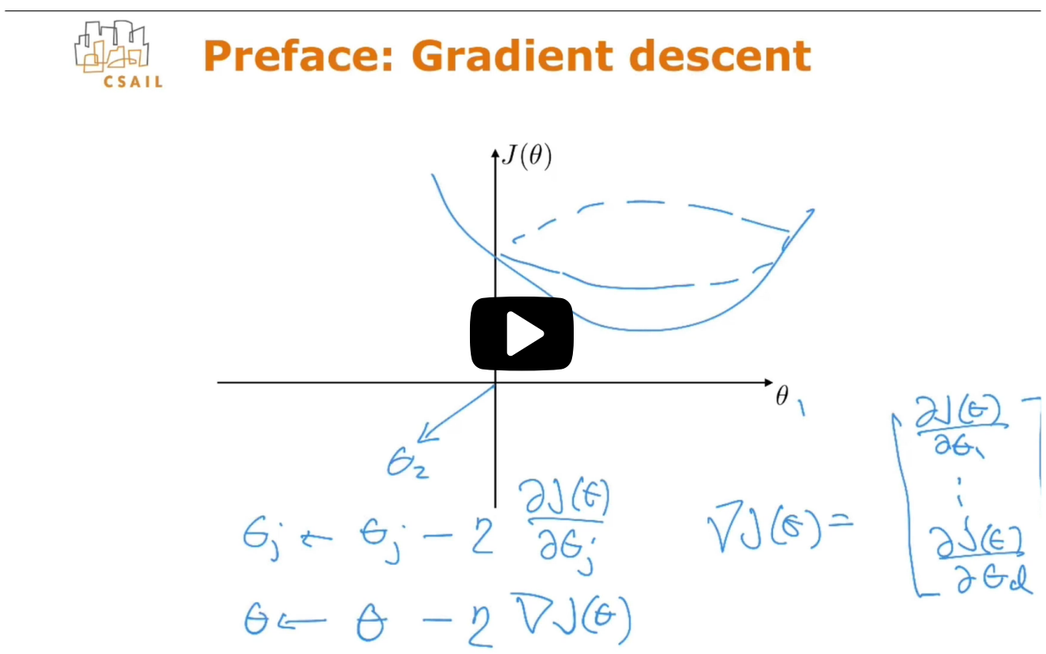
4. Gradient Descent
Gradient Descent



the gradient of the function with respect to 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.

▶ 4:19 / 4:19 Speed 1.50x

Video
Download video file

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

Gradient Descent: Geometrically Revisited

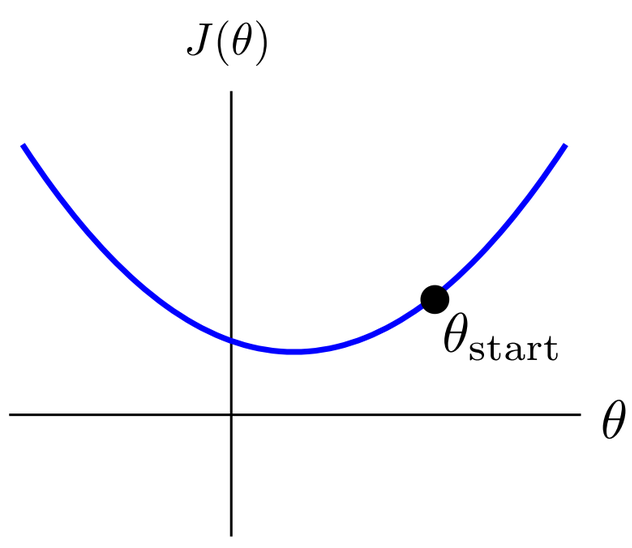
2/2 points (graded)
Assume $\theta \in \mathbb{R}$. Our goal is to find θ that minimizes

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

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 ▼

Gradient Hello, why don't we just find the gradient equal to zero to get to the minimum of the function? Thanks.	9
What is the difference of downwards and toward origin. I think move downwards is equaivalant to toward origin in this graph, am I wrong?	6
How can we define stepsize? How can we define stepsize?	1

☑	What is the relationship between Gradient Descent and Perceptron Algorithm? Since those two ways both used for updating parameter, what is the difference and relationship between them?	3
?	How do we determine that function is differentiable ? Wanted to ask this because it might be something I am not aware of. For an objective function J with theta in 'd' dimensions, how do we ensure that this function is differentia...	2
💬	[Staff] The word "downwards" was kind of misleading The way the sentence phrased "downwards" made it seem like the question was asking if we should decrease θ. Although the question is fine on itself, we had the graph figur...	4
?	What happens if we increase the stepsize η ? Not sure whether this problem is correctly defined. Regardless of the step size (no matter whether big (to some extent) or small) the update change is being smaller and small...	3
☑	Please confirm my understanding 1) Each derivative in a direction (theta1) is going to update, the parameter in that direction (theta1) but will not impact the parameters in other directions as those derivatives ...	6
☑	Gradient Descent: Geometrically Revisited: x missing superscript	2

Learn About Verified Certificates