4. Gradient Descent

Exercises due Sep 30, 2020 05:29 IST *Completed*

Gradient Descent

Start of transcript. Skip to the end.

So so far, we have seen how to qualitatively understand

the type of solutions that we get

when we vary the regularization parameter

and optimize with respect to theta and theta naught.

Now, we are going to talk about, actually, algorithms

Video
Download video file

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

Gradient Descent: Geometrically Revisited

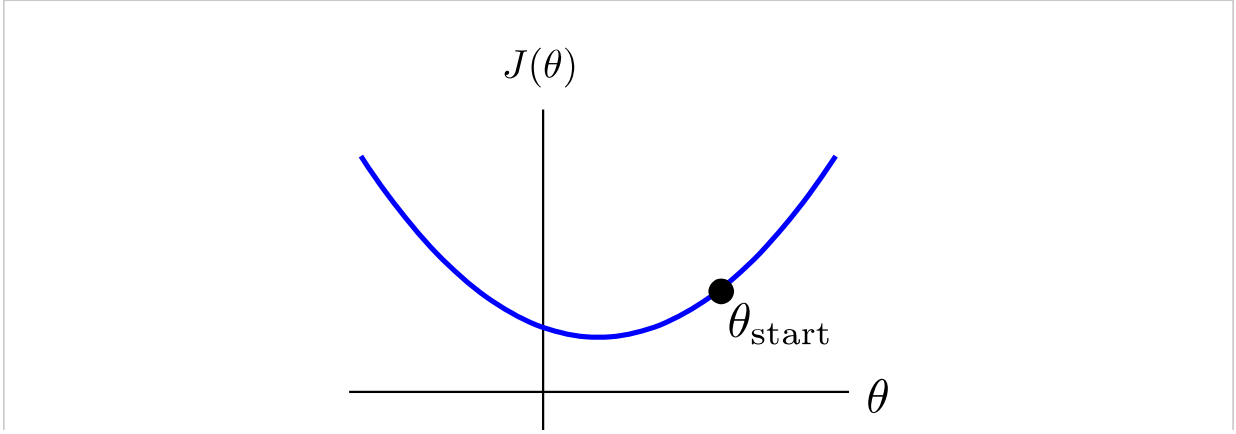
2/2 points (graded)
Assume $\theta \in \mathbb{R}$. Our goal is to find θ that minimizes

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

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

✔

Submit You have used 2 of 3 attempts

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

Blogpost explaining gradient descent Here's an excellent resource for those who are trying to gain a better understanding of the concept of the gradient descent: https://britt...	2
Towards/away from the origin? upwards/downwards? As the question statement says nothing about the step size, any of the answers could be correct. I know what the expected answer is, b...	5
Formulation of the second question I was confused by the proposed answers to second question, which i believe are misleading. The value of eta controls the size of the ste...	2
Minima of the objective function Can someone confirm my understanding - Since both the loss function and the generalization term are convex functions, their sum, the ...	1
Basic Math Question about θ_{start}	2
Gradient Descent: Geometrically Revisited While hinge loss is convex it is not differentiable, especially at margin boundaries. So how is step 2 justified ?	5

Previous Next