5. Gradient Based Approach

Learning Algorithm: Gradient Based Approach

0:00 / 0:00 1.50x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

True or False

0 points possible (ungraded)

Let $R_n(\theta)$ be the least squares criterion defined by

$$R_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \text{Loss}\left(y^{(t)} - \theta \cdot x^{(t)}\right).$$

Which of the following is true? Choose all those apply.

- [x] The least squares criterion $R_n(\theta)$ is a sum of functions, one per data point.
- [ ] Stochastic gradient descent is slower than gradient descent.
- [x] $\nabla_\theta R_n(\theta)$ is a sum of functions, one per data point.

Solution:

For every point, the loss is a function of θ, so the least squares criterion $R_n(\theta)$ is a sum of functions, one per data point, and this is what makes stochastic gradient descent possible. We want to do stochastic gradient descent because it is faster than gradient descent. Finally, because $R_n(\theta)$ is sum of functions, one per data point, $\nabla_\theta R_n(\theta)$ is also a sum of functions one per data point.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Add a Post

Hide Discussion

Topic: Unit 2 Nonlinear Classification, Linear regression, Collaborative Filtering (2 weeks):Lecture 5. Linear Regression / 5. Gradient Based Approach

Show all posts by recent activity

- Hint: averages count as sums — 1
  I hope I'm not giving away too much here, but for anyone who is struggling with this problem...
- ? [STAFF] Question is not exactly covered in lecture — 14
  Hi team, I don't think the lecture covers the concept of a Sum of Functions,
- Why that learning rate? — 7
- How to take derivative — 4
  In the video on 3:11 the professor takes the derivative using the chain rule. Can somebody ex...
- Intuition for Gradient Descent possibly more clear by rewriting gradient — 5

Learn About Verified Certificates