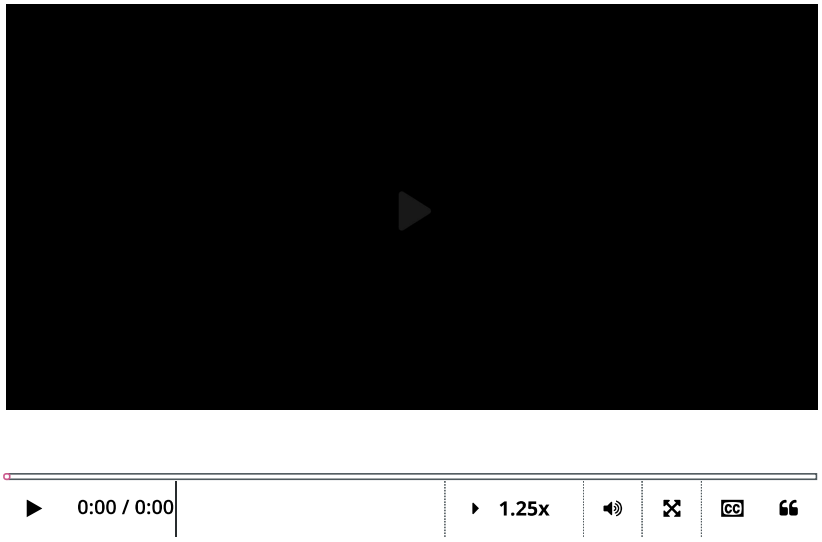
5. Gradient Based Approach

Learning Algorithm: Gradient Based Approach



Start of transcript. Skip to the end.

And the first algorithm that I want to demonstrate-- so we'll just close where we are, our objective. We are done, and now we are talking about finding algorithm. And as you can see here, we will start first with the gradient-based approach.

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.

- [ ] 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 3 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

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

Add a Post

Show all posts | by recent activity

?	[STAFF] One time extension Hello Staff, I've had very severe allergies that makes it very difficult to look at a screen due to light sensitivity and irritated eyes. I also ha...	2
💬	Question from previous section: Apparently, the empricial risk is just an approximation of the true risk (i.e. E[(f(x) −f^(x))2]). If we were able to learn the model by minimizing the true risk, which part of the error decomposition will become 0? @Staff/peers Please elaborate the answer for this question from the previous section. Question for you to think: Apparently, the empirici...	2
?	[Staff] Previous section. Empirical Risk and Model Performance Could someone explain the solution? We are not talking about complexity. Also we are talking about training and test data are drawn fr...	6
💬	[STAFF] Previous section 4 doesn't check complete Despite answering all questions the previous section doesn't turn green for me as well as the entire Lecture exercise 5 has not been ma...	3
?	Lec.5-5 loss term with factor 1/2 Hi, could someone clarify where the factor 1/2 comes from? thanks	4
☑	faster convergence with sgd? my intuition would tell me that using the gradient for the whole dataset would lead to faster convergence. why is my thinking wrong her...	7
?	Discussion form doesn't show up in the previous section Hi, Am I the only one not seeing the discussion forum section in the previous section? Thanks.	3
?	[STAFF] Sum of functions, one per data point? Rn(theta) is average of sum of function as we are dividing it by 'n' i.e is number of data points. Could you please confirm?	1
☑	How to understand self-correct note? The professor mentioned about one example that true target value is larger than prediction value, so "you are pushing your values in th...	2
?	Shouldn't the loss term be squared here? Least "square". If so, please correct. Perhaps a typo.	5

Previous | Next

Connect

/