5. Stochastic Gradient Descent

Exercises due Feb 22, 2023 08:59 -03 Completed

Stochastic Gradient Descent

▶ 0:00 / 0:00 1.0x

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

SGD and Hinge Loss

1/1 point (graded)

As we saw in the lecture above,

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

With stochastic gradient descent, we choose $i \in \{1, \ldots, n\}$ at random and update θ

$$\theta \leftarrow \theta - \eta \nabla_\theta \left[\text{Loss}_h\left(y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)\right) + \frac{\lambda}{2} \mid\mid \theta \mid\mid^2\right]$$

Submit You have used 2 of 3 attempts

Comparison with Perceptron

1/1 point (graded)
Observing the update step of SGD,

Which of the following is true?

- ( ) As in perceptron, is not updated when there is no mistake
- (•) Differently from perceptron, is updated even when there is no mistake

✓

Submit You have used 1 of 1 attempt

Discussion

Topic: Unit 1. Linear Classifiers and Generalizations (2 weeks):Lecture 4. Linear Classification and Generalization / 5. Stochastic Gradient Descent

Why does we update theta even if there is not mistake while doing stochastic gradient desce
I don't understand why does we update theta even if there is not mistake while doing stochastic gradient des

? Another varying learning rate parameter question
Does having the learning rate parameter sum to 1 as the number of iterations goes to infinity **guarantee** t