

### Machine Learning with Python-From Linear Models to Deep Learning

<u>Course</u> <u>Progress</u> <u>Dates</u> <u>Discussion</u> <u>Resources</u>

A Course / Unit 1. Linear Classifiers and Generali... / Lecture 4. Linear Classification



### 5. Stochastic Gradient Descent

 $\square$  Bookmark this page

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

#### **Stochastic Gradient Descent**



#### \_\_\_\_

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,

$$J\left( heta, heta_0
ight) = rac{1}{n}\sum_{i=1}^n \operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight) + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight) + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight) + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight) + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight) + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\mid\mid heta\mid\mid^2 = rac{1}{n}\sum_{i=1}^n \left[\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\left(\operatorname{Loss}_h\left(y^{(i)}\left( heta\cdot x^{(i)} + heta_0
ight)
ight] + rac{\lambda}{2}\left(\operatorname{Loss}_h\left(y^{(i)}\left( heta_0
ight)
ight] +$$

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

$$heta \leftarrow heta - \eta 
abla_{ heta} igl[ \operatorname{Loss}_h \left( y^{(i)} \left( heta \cdot x^{(i)} + heta_0 
ight) 
ight) + rac{\lambda}{2} \mid\mid heta \mid\mid^2 igr]$$

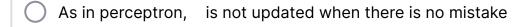
Submit

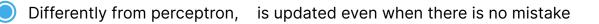
You have used 2 of 3 attempts

## Comparison with Perceptron

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

Which of the following is true?







Submit

You have used 1 of 1 attempt

Discussion < Previous Next >

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



# edX

**About** 

**Affiliates** 

edX for Business

Open edX

<u>Careers</u>

News

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

## Connect

Blog **Contact Us Help Center** Security Media Kit















© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 粤ICP备17044299号-2