4. Hinge Loss and Objective Function

Hinge Loss and Objective Function

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Hinge Loss Exercise 1

3/3 points (graded)

Compute the output of Hinge Loss function (as described in the video) for the following values:

$\text{Loss}_h(0) =$ 1 ✔ **Answer:** 1

$\text{Loss}_h(0.2) =$ 0.8 ✔ **Answer:** 0.8

$\text{Loss}_h(-10) =$ 11 ✔ **Answer:** 11

Solution:

$$\text{Loss}_h\left(z\right) = \begin{cases} 0 \text{ if } z >= 1 \\ 1 - z \text{ otherwise} \end{cases}$$

Submit

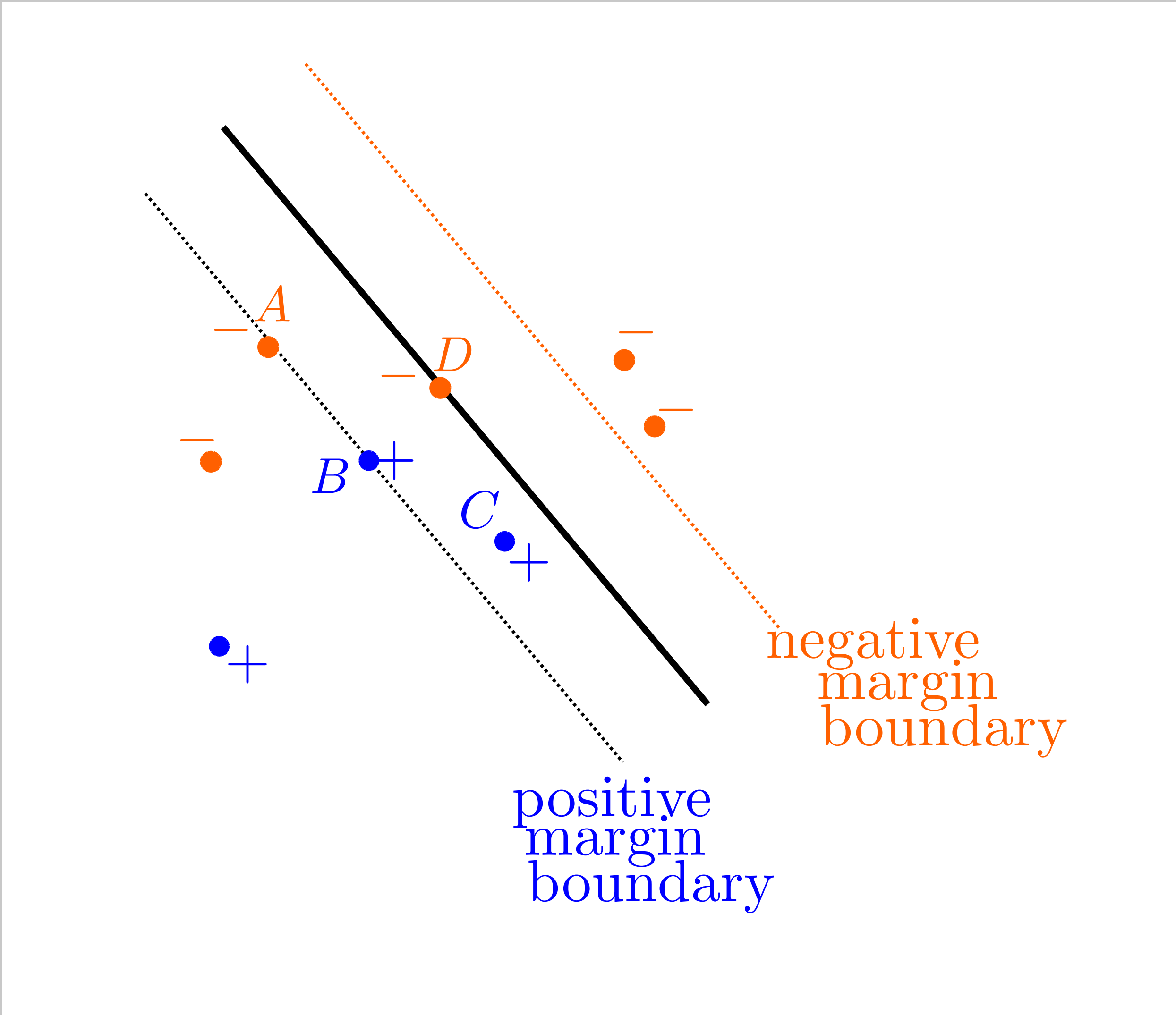
You have used 1 of 2 attempts

Answers are displayed within the problem

Hinge Loss Exercise 2

4/4 points (graded)

In a 2 dimensional space, there are points A, B, C, D as depicted below. Let $A = \left(x_a, y_a\right), B = \left(x_b, y_b\right), C = \left(x_c, y_c\right), D = \left(x_d, y_d\right)$.



What is the hinge loss of point A, $\text{Loss}_h\left(y^{(a)}\left(\theta \cdot x^{(a)}+\theta_0\right)\right)$?

- 0
- between 0 and 1
- 1
- 2 (selected)

What is the hinge loss of point B, $\text{Loss}_h\left(y^{(b)}\left(\theta \cdot x^{(b)}+\theta_0\right)\right)$?

- (•) 0
- ( ) between 0 and 1
- ( ) 1

✓

What is the hinge loss of point C, $\text{Loss}_h\left(y^{(c)}\left(\theta \cdot x^{(c)} + \theta_0\right)\right)$?

- ( ) 0
- (•) between 0 and 1
- ( ) 1

✓

What is the hinge loss of point D, $\text{Loss}_h\left(y^{(d)}\left(\theta \cdot x^{(d)} + \theta_0\right)\right)$?

- ( ) 0
- ( ) between 0 and 1
- (•) 1

✓

Solution:

A is on the positive margin boundary but with the label -1, so

$$y^{(a)}\left(\theta \cdot x^{(a)} + \theta_0\right) = -1.$$

Thus its hinge loss is 2. B is on the positive margin boundary and with the label $+1$, so

$$= y^{(b)}\left(\theta \cdot x^{(b)} + \theta_0\right) = 1.$$

Thus its hinge loss is 0. C lies between the decision boundary and the margin boundary. Thus

$$1 > y^{(c)}\left(\theta \cdot x^{(c)} + \theta_0\right) > 0.$$

Thus C's hinge loss is between 0 and 1. Similarly, because D is on the decision boundary,

$$y^{(d)}\left(\theta \cdot x^{(d)} + \theta_0\right) = 0.$$

Thus its hinge loss is 1.**Loss functions tell you in general how bad the prediction is.** The Hinge Loss tells us how undesirable a training example is, with regard to the margin and the correctness of its classification.

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Regularization

1/1 point (graded)

Remember that for points (x, y) on the boundary margin, the distance from the decision boundary to (x, y) is $\frac{1}{||\theta||}$. Thus

$$y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) = 1.$$

And

$$\frac{y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)}{||\ \theta\ ||} = \frac{1}{||\ \theta\ ||}.$$

Now our goal is to maximize the margin, that is to maximize $\frac{1}{||\theta||}$. Which of the following is **NOT** equivalent to maximizing $\frac{1}{||\theta||}$?

- maximizing $\frac{1}{||\theta||^2}$
- minimizing $||\ \theta\ ||$
- maximizing $\sqrt{||\ \theta\ ||}$ (selected, correct)

Solution:

Maximizing $\frac{1}{||\theta||}$ is equivalent to maximizing $\frac{1}{||\theta||^2}$. It is also equivalent to minimizing $||\ \theta\ ||$.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Objective

1/1 point (graded)

Remember that our objective is given as

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

Our goal is to minimize this objective J. Now, which of the following is true if we have a large λ?

- (selected) We put more importance on maximizing the margin than minimizing errors
- We put more importance on minimizing the margin than minimizing errors
- We put more importance on maximizing the margin than maximizing errors
- We put more importance on minimizing the margin than maximizing errors

✔

Solution:

Remember that the first term

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

corresponds to the sum of hinge losses on each training example, and the second term

$$\frac{\lambda}{2} \mid\mid \theta \mid\mid^2$$

corresponds to maximizing the margin. If we increase λ, we put more weight on maximizing the margin than minimizing the sum of losses.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Lecture 3 Hinge loss, Margin boundaries and Regularization / 4. Hinge Loss and Objective Function

Add a Post

Show all posts by recent activity

- why the distance from the decision boundary to (x,y) must be $1/||\theta||$? 5
  why the distance from the decision boundary to (x,y) must be $1/||\theta||$? Can the distance be 2...
- [Staff] Can we get annotated slides? 6
  Thanks for posting the clean slides. Much of the material in the lectures is in form of annotati...
- why do we need pos margin boundary and neg margin boundary? 5
  we already have decision boundary,why do we need pos margin boundary and neg margin b...
- Hinge Loss definition 3
- Hinge Loss Exercise 1 6
  What is it that I am missing in this exercise? It tells you to "Compute the output of Hinge Loss ...

Learn About Verified Certificates