1. Objective

Bookmark this page

Hinge loss, Margin boundaries, and Regularization

At the end of this lecture, you will be able to

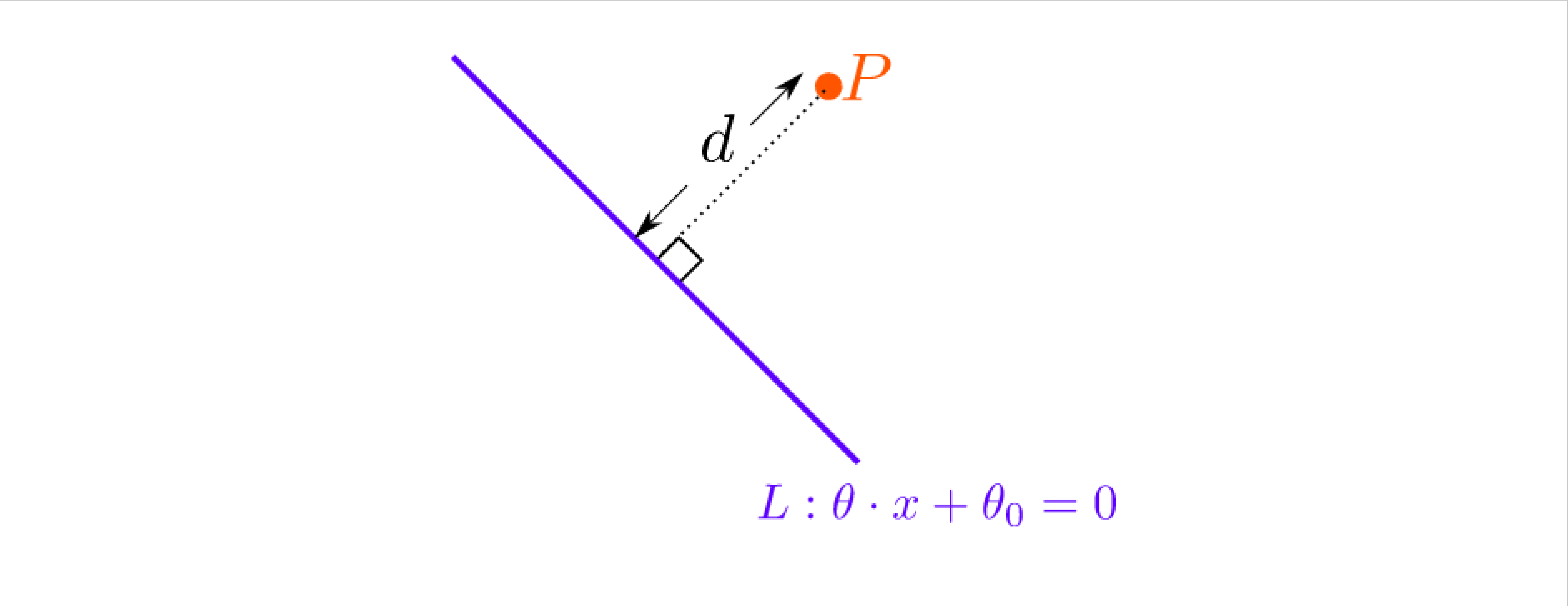
- understand the need for maximizing the margin
- pose linear classification as an optimization problem
- understand hinge loss, margin boundaries and regularization

Consider a line L in $\mathbb{R}^2$ given by the equation

$$L : \theta \cdot x + \theta_0 = 0$$

where θ is a vector normal to the line L. Let the point P be the endpoint of a vector x_0 (so the coordinates of P equal the components of x_0).

What is the the shortest distance d between the line L and the point P? Express d in terms of θ, θ_0, x, x_0.



$d =$

- ( ) $\frac{|\theta \cdot x + \theta_0|}{||\ \theta\ ||}$
- (•) $\frac{|\theta \cdot x_0 + \theta_0|}{||\ \theta\ ||}$
- ( ) $\frac{|\theta \cdot \theta_0 + \theta_0|}{||\ \theta\ ||}$
- ( ) $|\theta \cdot x_0 + \theta_0|$

✔

Solution:

If there is no offset θ_0, The distance d is the projection from x_0 to θ, which is $\frac{|x_0 \cdot \theta|}{||\theta||}$(definition of projection). With the offset θ_0 added, d is $\frac{|x_0 \cdot \theta + \theta_0|}{||\theta||}$. Thus the distance from a $L : \theta \cdot x + \theta_0 = 0$ to the point $P = x_0$ is given by $\frac{|\theta \cdot x_0 + \theta_0|}{||\theta||}$.

Show answer

Submit You have used 2 of 3 attempts

The **decision boundary** is the set of points x which satisfy

$$\theta \cdot x + \theta_0 = 0.$$

The **Margin Boundary** is the set of points x which satisfy

$$\theta \cdot x + \theta_0 = \pm 1.$$

So, the distance from the decision boundary to the margin boundary is $\dfrac{1}{\|\theta\|}$.

Margin Boundary 1

1/1 point (graded)

As explained in the lecture video, margin boundary is the set of points (x, y) at which the distance from the decision boundary to (x, y) is $\frac{1}{||\theta||}$. Now, what is the value of $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)$ for a correctly classified point $(x^{(i)}, y^{(i)})$ on the margin boundary?

1

Answer: 1

Solution:

Solution:

From the previous problem, we know that the distance from a line $L : \theta x + \theta_0 = 0$ to $P = (x_0)$ is given by $\frac{||\theta x_0 + \theta_0||}{||\theta||}$.
Because we know that the distance from the decision boundary to (x, y) is $\frac{1}{||\theta||}$,

$$|| \ \theta x_0 + \theta_0 \ || = 1$$

. Thus,

$$|| \ \theta x_0 + \theta_0 \ || = y^{(i)} \left(\theta \cdot x^{(i)} + \theta_0\right) = 1$$

Show answer

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Margin Boundary 2

1/1 point (graded)

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:

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

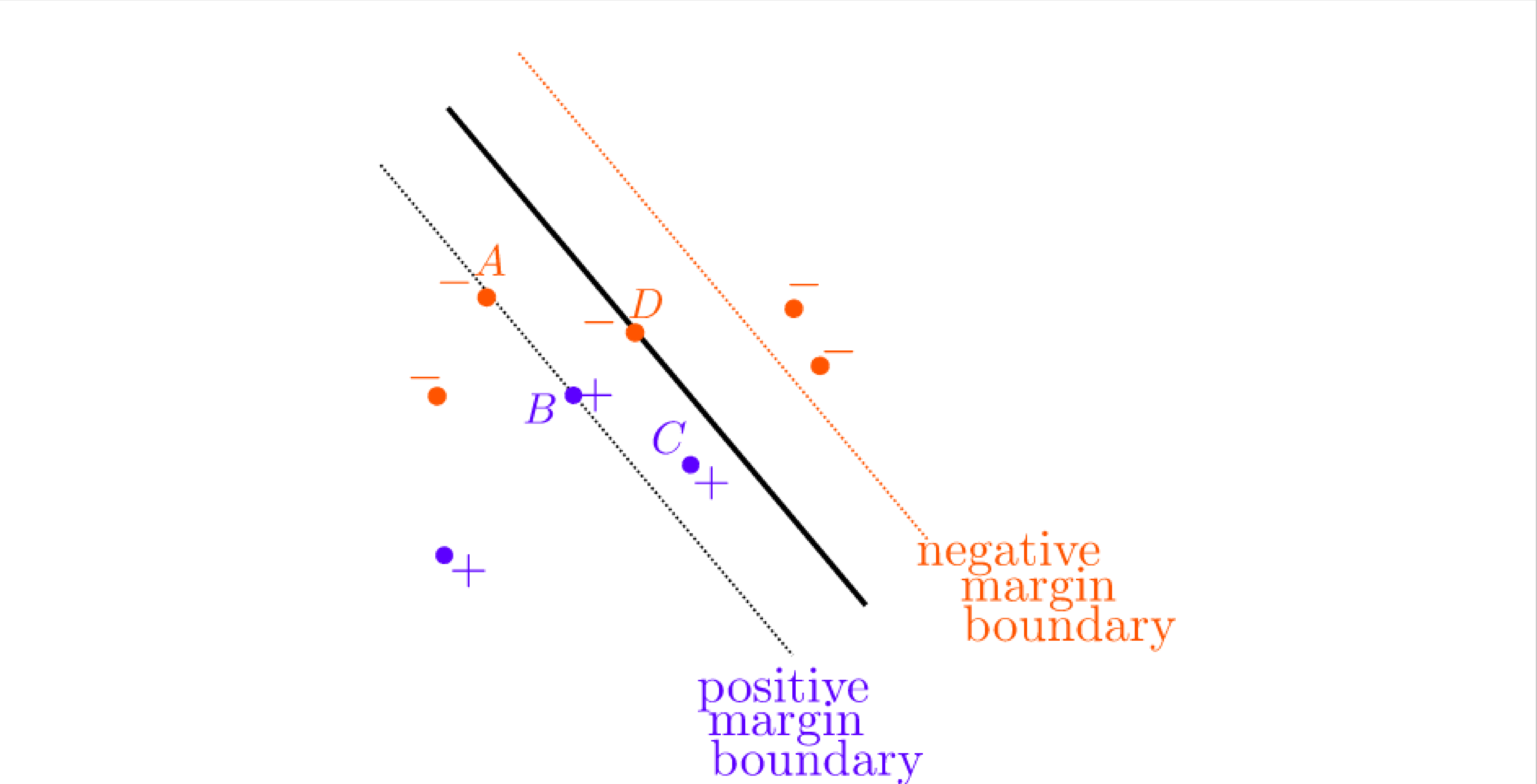
Show answer

Submit You have used 2 of 2 attempts

Hinge Loss Exercise 2

4/4 points (graded)

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



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

✓

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, $\mathrm{Loss}_h\left(y^{(d)}\left(\theta \cdot x^{(d)} + \theta_0\right)\right)$?

- ◯ 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.

Show answer

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 $\sqrt{||\ \theta\ ||}$

✔

Solution:

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

Show answer

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

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

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

- [x] 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

$$\frac{\lambda}{2}\,||\,\theta\,||^2$$

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