

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#). ✕



[Unit 1 Linear Classifiers and](#)
[Course](#) > [Generalizations \(2 weeks\)](#) > [Homework 2](#) > 1. Linear Support Vector Machines

1. Linear Support Vector Machines

In this problem, we will investigate minimizing the training objective for a Support Vector Machine (with margin loss).

The training objective for the Support Vector Machine (with margin loss) can be seen as optimizing a balance between the average hinge loss over the examples and a regularization term that tries to keep the parameters small (increase the margin). This balance is set by the regularization parameter $\lambda > 0$. Here we only consider the case without the offset parameter θ_0 (setting it to zero) so that the training objective is given by

$$\left[\frac{1}{n} \sum_{i=1}^n \text{Loss}_h(y^{(i)} \theta \cdot x^{(i)}) \right] + \frac{\lambda}{2} \|\theta\|^2 = \frac{1}{n} \sum_{i=1}^n \left[\text{Loss}_h(y^{(i)} \theta \cdot x^{(i)}) + \frac{\lambda}{2} \|\theta\|^2 \right] \quad (3.3)$$

where the hinge loss is given by

$$\text{Loss}_h(y(\theta \cdot x)) = \max\{0, 1 - y(\theta \cdot x)\}$$

$$\hat{\theta} = \text{Argmin}_{\theta} [\text{Loss}_h(y\theta \cdot x) + \frac{\lambda}{2} \|\theta\|^2] \quad (3.4)$$

Note: For all of the exercises on this page, assume that $n = 1$ where n is the number of training examples and $x = x^{(1)}$ and $y = y^{(1)}$.

Minimizing Loss - Case 1

1/1 point (graded)

In this question, suppose that $\text{Loss}_h(y(\hat{\theta} \cdot x)) > 0$. Under this hypothesis, solve for optimisation problem and express $\hat{\theta}$ in terms of x , y and λ

$y \cdot x / \lambda$

✓ Answer: $x \cdot y / \lambda$

$\frac{y \cdot x}{\lambda}$

STANDARD NOTATION

Solution:

$$\hat{\theta} = \text{Argmin}_{\theta} [\text{Loss}_h(y\theta \cdot x) + \frac{\lambda}{2} \|\theta\|^2]$$

The above loss can be minimized by solving for the following equation

$$0 = \nabla_{\theta} [\text{Loss}_h(y(\theta \cdot x))] + \nabla_{\theta} [\frac{\lambda}{2} \|\theta\|^2]$$

Given that

$$\begin{aligned} \text{Loss}_h(y(\hat{\theta} \cdot x)) &> 0 \\ \text{Loss}_h(y(\hat{\theta} \cdot x)) &= \max\{0, 1 - y(\theta \cdot x)\} \\ \text{Loss}_h(y(\hat{\theta} \cdot x)) &= 1 - y(\theta \cdot x) \\ \nabla_{\theta} [\text{Loss}_h(y(\theta \cdot x))] &= -yx \end{aligned}$$

Plugging this back in the previous equation, we get:

$$0 = \lambda \hat{\theta} - yx$$

$$\hat{\theta} = \frac{1}{\lambda} yx$$

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Minimizing Loss - Numerical Example (1)

2/2 points (graded)

Consider minimizing the above objective function for the following numerical example:

$$\lambda = 0.5, y = 1, x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Note that this is a classification problem where points lie on a two dimensional space. Hence $\hat{\theta}$ would be a two dimensional vector.

Let $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2]$, where $\hat{\theta}_1, \hat{\theta}_2$ are the first and second components of $\hat{\theta}$ respectively.

Solve for $\hat{\theta}_1, \hat{\theta}_2$.

Hint: For the above example, show that $\text{Loss}_h(y(\hat{\theta} \cdot x)) \leq 0$

$$\hat{\theta}_1 =$$

✓ Answer: 1.0

 $\hat{\theta}_2 =$

✓ Answer: 0.0

Solution:

First note that for this example $Loss_h(y(\theta \cdot x)) \leq 0$.

To show this we use proof by contradiction.

Suppose $Loss_h(y(\theta \cdot x)) > 0$:

From the previous problem, we know that under this condition, $\hat{\theta} = \frac{yx}{\lambda}$

For this example, $\hat{\theta} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

For this value of $\hat{\theta}$, we see that $1 - (y(\theta \cdot x)) = 1 - 2 = -1 < 0$ contradicting our original assumption.

Hence, $Loss_h(y(\theta \cdot x)) \leq 0$, which implies that $y(\theta \cdot x) \geq 1$.

We are left with minimizing $\frac{\lambda}{2} \|\theta\|^2$ under the constraint $y(\theta \cdot x) \geq 1$.

The geometry of the problem implies that in fact, $y(\theta \cdot x) = 1$.

That is, $1 - (\hat{\theta}_1 * 1 + \hat{\theta}_2 * 0) = 0$ implying that $\hat{\theta}_1 = 1$.

Then, to minimize $\|\theta\|$, $\hat{\theta}_2 = 0$.

Therefore $\hat{\theta} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

In fact, we can show that $\hat{\theta} = \frac{x}{y\|x\|^2}$. Looking back at the previous question, the solution of the optimization is then necessarily of the form $\hat{\theta} = \eta y x$ for some real $\eta > 0$.

Submit

You have used 2 of 3 attempts

i Answers are displayed within the problem

Minimizing Loss - Numerical Example (2)

1.0/1 point (graded)

Now, let $\hat{\theta} = \hat{\theta}(\lambda)$ be the solution as a function of λ .

For what value of $\|x\|^2$, the training example (x, y) will be misclassified by $\hat{\theta}(\lambda)$?

$\|x\|^2 =$

0

✓ Answer: 0

Solution:

For a point to be considered misclassified

$$y\hat{\theta} \cdot x \leq 0$$

The above condition implies that the hinge loss is greater than zero. From above problems, we know that under this condition,

$$\hat{\theta} = \frac{yx}{\lambda}$$

$$y\hat{\theta} \cdot x = \frac{y^2 \|x\|^2}{\lambda} \leq 0$$

All terms of the product are non-negative, making it impossible to be < 0 . But if $\|x\| = 0$, the product can be 0.

Hence $\|x\|^2 = 0$

Submit

You have used 1 of 3 attempts

i Answers are displayed within the problem

Discussion





Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Homework 2 / 1. Linear Support Vector Machines

[Hide Discussion](#)
[Add a Post](#)

Show all posts ▼

by recent activity ▼

- | | |
|--|-----------|
| <input checked="" type="checkbox"/> [STAFF] More explanation needed for answer to 2nd question
Particularly: "In fact, we can show that $\theta^* = xy \ (\ x\)^2$. Looking back at the previous question, the solution of the optimization is then necessaril... | 4 |
| [Staff] Bad typing in Numerical Example (1). | 2 |
| So the takeaway is...?
I got all the green ticks for section 1 in homework 2. However, I don't know what I should have learned from this exercise of mathematical deduc... | 8 |
| [Staff] Minimizing Loss - Numerical Example (2) - Curious wording
Community TA | 11 |
| Minimizing Loss - Case 1; Question about model solution | 2 |
| Overall impression about this course is bad, so far.
Almost all the problems require further clarification from the TAs and reading through the discussion board. Frustrating and time consuming ex... | 9 |
| Minimizing Loss - Case 1
Hey, I have tried to submit this answer many times but it doesn't submit. Where is the problem? $(x^* y) / -\lambda$ | 2 |
| <input checked="" type="checkbox"/> Minimizing Loss - Numerical Example (1) | 4 new_ 18 |
| <input checked="" type="checkbox"/> Are there more solutions to Minimizing Loss - Numerical Example (2)?
I got the green tick but the wording of the problem leads me to believe my solution might be lacking something. Is the wording misleading? | 6 |
| <input checked="" type="checkbox"/> Minimizing Loss - Numerical Example (1) Hint: how Loss could be negative? | 21 |

-  [Staff] The project due date appears as 4th july and the homework/Excercises for today. 1
Hi, Since, we had technical problems with the site last week, Could you, please, set both the homework and exercises to the same date as the pro...
-  Hint for Minimizing Loss - Numerical Example (1) 6
For all not soooo Math savvy as myself: I turned almost crazy trying to solve the minimizing equation with two variables under constraints. Then I...
-  [STAFF] Is it possible to reset attempts 3
For question 1 I made the mistake of not calculating the correct derivative because I forgot to derivate the 1 in the equation. You can see this if y...
-  Question about optimization case 1 5

© All Rights Reserved