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

$$\left[\frac{1}{n}\sum_{i=1}^{n} Loss_h\left(y^{(i)}\,\theta\cdot x^{(i)}\,\right)\right]+\frac{\lambda}{2}\|\theta\|^2=\frac{1}{n}\sum_{i=1}^{n}\left[Loss_h\left(y^{(i)}\,\theta\cdot x^{(i)}\,\right)+\frac{\lambda}{2}\|\theta\|^2\right] \tag{3.3}$$

where the hinge loss is given by

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

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

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

y*x/lambda

Answer: x*y/lambda

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

STANDARD NOTATION

Solution:

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

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

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

Given that

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

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

Answers are displayed within the problem

Minimizing Loss - Numerical Example (1)

2/2 points (graded)

Consider minimizing the above objective fuction 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 $\hat{\theta} = \left[\hat{\theta}_1, \hat{\theta}_2 \right]$, 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 $\text{Loss}_h \left(y \left(\hat{\theta} \cdot x \right) \right) \leq 0$

$\hat{\theta}_1 =$

1

Answer: 1.0

$\hat{\theta}_2 =$

0

Answer: 0.0

Solution:

First note that for this example $Loss_h\left(y\left(\theta \cdot x\right)\right) \leq 0$.
To show this we use proof by contradiction.

Suppose $Loss_h\left(y\left(\theta \cdot x\right)\right) > 0$:
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 $1 - \left(y\left(\theta \cdot x\right)\right) = 1 - 2 = -1 < 0$ contradicting our original assumption.

Hence, $\mathrm{Loss}_h\left(y\left(\theta \cdot x\right)\right) \leq 0$, which implies that $y\left(\theta \cdot x\right) \geq 1$.

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

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

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

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

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

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 2 of 3 attempts

Answers are displayed within the problem

Discussion

Add a Post

Hide Discussion

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

Show all posts

by recent activity

- Two homeworks in one week is strange.
  I had to go to a different country for 10 days, and I didn't have much access to the internet, so I missed... 2
- Minimizing Loss - Numerical Example (1) vs. Case 1 [staff]
  Not clear about what happens here. I gave the correct answer in Case 1, tried to apply the same reaso... 7
- ? Norm x^2
  Numerical or with lambda? 2
- ? What`s theta_hat?
  Can anyone help me understand what theta_hat is? I probably missed it in the lectures... 4
- minimizing case 1 _theta not accepted in answer
  I am so confused. From the lectures it appears that argmin of theta_hat, is a derivative of the loss funct... 3
- Minimizing Loss - Numerical Example (1) 8 new 11
- Minimizing Loss - Numerical Example (2): need a hint 18 new 25
- Numerical Example 1
  After having spent some time on this question, I thought I could give a quick hint to anyone struggling (... 7
- Need Help, Clarification, Anything (Numerical Example 2)
  I'm absolutely stuck on this question for over 3,4 hours. I think I understand the role of \lambda and th... 11
- suggestion
  I'd like to suggest that you abolish the word "above". E.g. instead of saying "the above example" or "the... 4
- ? [Cace 1] theta instaead of theta_hat 4
- optimizing perceptron
  FYI, I come from a Computer Science background. given this version of the perceptron algorithm with ... 1
- A recitation should have been included to review constrained and unconstrained loss minimization concepts.
  Just wanted to share that I was going really fine up till this homework and suddenly got dejected and d... 12 new

Learn About Verified Certificates