4. Linear Support Vector Machines

Bookmark this page

Homework due Sep 30, 2020 05:29 IST *Completed*

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

$$\left[\frac{1}{n}\sum_{i=1}^{n} Loss_h\,(y^{(i)}\,\theta\cdot x^{(i)}\,)\right] + \frac{\lambda}{2}\|\theta\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left[Loss_h\,(y^{(i)}\,\theta\cdot x^{(i)}\,) + \frac{\lambda}{2}\|\theta\|^2\right] \tag{4.3}$$

where the hinge loss is given by

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

$$\hat{\theta} = \text{Argmin}_\theta\,[\text{Loss}_h\,(y\,\theta\cdot x\,) + \frac{\lambda}{2}\|\theta\|^2] \tag{4.4}$$

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*x/lambda

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

✔ **Answer:** x*y/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

Answers are displayed within the problem

Minimizing Loss - Numerical Example (1)

1/2 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 $\text{Loss}_h\left(y\left(\hat{\theta} \cdot x\right)\right) \leq 0$

$\hat{\theta}_1 =$

0.25 ✖ **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, $\text{Loss}_h\left(y\left(\theta \cdot x\right)\right) \leq 0$, 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, $1-(\hat{\theta}_1 * 1+\hat{\theta}_2 * 0)=0$ 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 $t\hat{heta}=\frac{x}{y\|(\|x)^2}$. 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 1 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 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Homework 1 / 4. Linear Support Vector Machines