MITx: 6.86x
Machine Learning with Python-From Linear Models to Deep Learning

Help sandipan_dey

Course Discussion Progress Resources

Discussion

Cool7

Community TA, Student

5 discussions started 64 comments

Show all posts ▼ by recent activity ▼

?	Loss Function and decision boundaries -- Request for feedback	9
?	Calculating the step size? [Spoil Alert] Hello First of all, I need to say think you for this excellent course. Much appreciated. As for quesstion 2. It took me around 2 hours to guess the an correct answer. Here is my guess: The new theta gets updated by: θ(k+1) = θ...	6
	How can PA algorithm use zero-one loss? Community TA	15
	What's the superscript k in last option indicate?	7
	get_order(feature_matrix.shape[0]) Can someone let me know where the get_order() function is implemented? The one implemented in the project1.py is not the one called here, right?	5
	Question of Support Vectors Community TA	11
	Derivation of the error decomposition. (spoiler alert: don't look if you try to work it out yourself) Community TA	1
	Complaint based on my 27 years of C++ programming in regression. If you use a large enough training set, say 1000 times more samples than weights, which is well distributed, then indeed zero risk function should practically mean zero error also on the training set but this question is highl...	3
	[Staff] possible issue with the get_order() results for 'Full Perceptron Problem' I submitted my code multiple times and it failed the tests. When I was looking at the output that it produced, I realise that the ordering produced by the get_order() is different on my machine than the grader. I used the foll...	15

Typo in top description?

Shouldn't this equation: θ(k+1) = θ(k) + ηyx be θ(k+1) = θ(k) - ηyx Considering new θ in gradient decent is = θ - η*(Gradient) when trying to minimize loss

3

[Staff]True or False

The equation misses the equal sign. It seems the question come before the relevant lecture.

Community TA

3

Load more

All Posts

How can PA algorithm use zero-one loss?

question posted 13 days ago by **Cool7** (Community TA)

EDIT: Resolved, see explanation below.

I'm not sure if I get what question "Loss functions and decision boundaries" ask. For me, let PA use zero-one loss doesn't seem work.

PA(without offset)'s objective is finding θ that minimizes

$$\frac{\lambda}{2}\left\|\theta-\theta^{(k)}\right\|^2+\text{Loss}\left(y\theta\cdot x\right)$$

And if we are using zero-one loss, and according to the data provided in question,

$$Loss\left(y\theta\cdot x\right)=1$$

Thus, we are now look at this objective function:

$$\frac{\lambda}{2}\left\|\theta-\theta^{(k)}\right\|^2+1$$

To minimize it, result is always $\theta=\theta^{(k)}$, independent of λ used.

What is going on here?

EDIT: Thanks for @mrBB point out my issue. I will do a brief explanation here.

When optimize P-A with 0/1 loss. Object function is
$\frac{\lambda}{2}\left\|\theta-\theta^{(k)}\right\|^2+\text{Loss}_{0/1}\left(y\theta\cdot x\right)$

$$\text{Loss}_{0/1}\left(y\theta\cdot x\right)=\begin{cases}0, & y\theta\cdot x>0\\ 1, & y\theta\cdot x<=0\end{cases}$$

To optimize this function when λ is big, value of $Loss_{0/1}$(either 0 or 1) become not so important. Result will be similar to optimize $\frac{\lambda}{2}\left\|\theta-\theta^{(k)}\right\|^2$ alone.

When λ is small(enough), value of $Loss_{0/1}$(either 0 or 1) become significant. Result will be push $Loss_{0/1}$ to 0 then optimize $\frac{\lambda}{2}\left\|\theta-\theta^{(k)}\right\|^2$ in this range(while $Loss_{0/1}=0$).

This post is visible to everyone.

mrBB
4 days ago - marked as answer 4 days ago by **Cool7**

@Cool7 I'm a bit late and perhaps you already figured it out. Why do you assume 0-1 loss will always be 1. Any θ that correctly classifies the data point in question, would result in zero 0-1loss, wouldn't it? (Per the definition of 0-1 loss). For example in graphs (a) and (c) we see the point would be correctly classified after the update and that it therefore would incur a 0-1 loss of zero. On the other hand in graphs (b) and (d) the point remains misclassifies and we would incur a 0-1 loss of 1.

Show Comments (2)

Add a Response

7 other responses

BrendanWood
13 days ago

+1 I've got the same question ... I don't see how options 3 and 4 would be different from each other and therefore can't uniquely classify the four plots.

Add a comment

david301
13 days ago

It doesn't pay to update θ when the loss term remains 1, but you might have the opportunity to change that loss term to 0.

Not sure what did you mean, but doesn't matter loss is 0 or 1, that θ is not going to budge.

What kind machine learning algorithm is this "not learning at all type"?

posted 13 days ago by **Cool7**

but with a small lambda if the loss is able to change to 0 this can cancel out the non favorable $||\theta - \theta^k||$ thus leading to a change

posted 12 days ago by **Jet_elephant**

Add a comment

BrendanWood
12 days ago

Oh, I was initially assuming $Loss\left(y\theta \cdot x\right) = 1$ as well, but that should actually read $Loss\left(y\theta^{(k)} \cdot x\right) = 1$. Hope this helps.

can you please explain what you mean by "with small lambda, loss is 0 " ?

posted 12 days ago by nr7116

@BrendanWood: Thank you for hint but how this prevents coclusion of OP regarding $\theta = \theta^{(k)}$ independant of λ? Should I take my derivative (gradient) with respect to $\theta^{(k)}$?

posted 5 days ago by sharov_am

Add a comment

paladin1410
7 days ago

Any one has an answer for this problem? How zero-one loss function can be used to update the weight?

Add a comment

chan127ck
6 days ago

I have same question. Can anyone explain in more details?

Add a comment

glanz
4 days ago

Just a few observations. The zero-one loss function seems to be just the 'reflected' Heaviside function $H\left(z\right)$. Whos derivative and antiderivative are $\delta\left(z\right)$ and the ramp function $R\left(z\right) := \max\{0, z\}$, respectively. What is more, the latter one is the 'reflected' hinge loss.

Add a comment

zitong0103
3 days ago

+1. Thank you Cool7.

Add a comment

Showing all responses

Add a response:

Preview

Submit