Why does the L2 norm loss have a unique solution and the L1 norm loss have possibly multiple solutions?

Ask Question

Asked 3 years ago Active 1 year, 6 months ago Viewed 5k times

21

10

http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/

If you look at the top of this post, the writer mentions that L2 norm has a unique solution and L1 norm has possibly many solutions. I understand this in terms of regularization, but not in terms of using L1 norm or L2 norm in the loss function.

If you look at graphs of functions of scalar x (x^2 and |x|), you can easily see both have one unique solution.

regression lasso regularization

Share Cite Improve this question Follow

edited Aug 21 '18 at 2:54
Glen_b
250k 28 529 909

asked Aug 21 '18 at 1:58
user3180
511 3 14

2 "fnx"? ... Please edit to make this clearer. Do you mean "functions"? – Glen_b Aug 21 '18 at 2:04

Add a comment

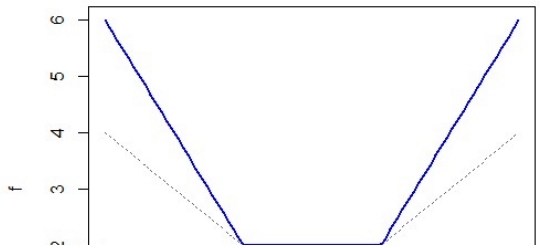
2 Answers

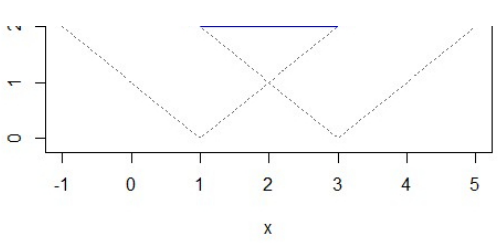
Active Oldest Votes

30

Let's consider a one-dimensional problem for the simplest possible exposition. (Higher dimensional cases have similar properties.)

While both $|x-\mu|$ and $(x-\mu)^2$ each have a unique minimum, $\sum_i |x_i-\mu|$ (a sum of absolute value functions with different x-offsets) often doesn't. Consider $x_1 = 1$ and $x_2 = 3$:





(NB in spite of the label on the x-axis, this is really a function of μ; I should have modified the label but I'll just leave it as is)

In higher dimensions, you can get regions of constant minimum with the L_1-norm. There's an example in the case of fitting lines here.

Sums of quadratics are still quadratic, so $\sum_i (x_i - \mu)^2 = n(\bar{x} - \mu)^2 + k(\mathbf{x})$ will have a unique solution. In higher dimensions (multiple regression say) the quadratic problem may not automatically have a unique minimum -- you may have multicollinearity leading to a lower-dimensional ridge in the negative of the loss in the parameter space; that's a somewhat different issue than the one presented here.

A warning. The page you link to claims that L_1-norm regression is robust. I'd have to say I don't completely agree. It's robust against large deviations in the y-direction, *as long as they aren't influential points* (discrepant in x-space). It can be arbitrarily-badly screwed up by even a single influential outlier. There's an example here.

Since (outside some specific circumstances) you don't usually have any such guarantee of no highly influential observations, I wouldn't call L1-regression robust.

R code for plot:

```
fi <- function(x,i=0) abs(x-i)
f <- function(x) fi(x,1)+fi(x,3)
plot(f,-1,5,ylim=c(0,6),col="blue",lwd=2)
curve(fi(x,1),-1,5,lty=3,col="dimgrey",add=TRUE)
curve(fi(x,3),-1,5,lty=3,col="dimgrey",add=TRUE)
```

Share Cite Improve this answer Follow edited Feb 11 '20 at 0:16

answered Aug 21 '18 at 2:13
Glen_b
250k 28 529 909

This is great. What software did you use to make the graph? – user3180 Aug 21 '18 at 2:20

2 R. This is just done in base graphics. I have added the code to the end of my answer. – Glen_b Aug 21 '18 at 2:23

1 Woah, never realised you could supply a function to `plot`. Mind is blown. – JAD Aug 21 '18 at 6:52

Are you talking about local minimums in this post or global minimum? When you say there is a unique minimum – user3180 Dec 7 '20 at 4:56

Add a comment

6

Minimizing the L2 loss corresponds to calculating the arithmetic mean, which is unambiguous, while minimizing the L1 loss corresponds to calculating the median, which is ambiguous if an even number of elements are included in the median calculation (see Central tendency: Solutions to variational problems).

Share Cite Improve this answer Follow edited Aug 22 '18 at 8:49

answered Aug 22 '18 at 8:02
HelloGoodbye
484 4 10

Add a comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email

Required, but never shown

Post Your Answer *By clicking "Post Your Answer", you agree to our terms of service, privacy policy and cookie policy*

Not the answer you're looking for? Browse other questions tagged regression lasso regularization or ask your own question.

Why does Google Maps say that driving from San Francisco to Seattle via Portland is 30 minutes faster than driving from San Francisco to Seattle?

Would interstellar space piracy be possible under these conditions?

If someone stops a potential murder by subduing a gunman, what would be the possible consequences?

Question feed

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email

Required, but never shown