Why is the geometric median called the L_1 estimator?

Ask Question

Asked 6 years, 11 months ago Active 6 years, 6 months ago Viewed 3k times

9

My question is simply, why is the geometric median called the L_1 estimator? This always reminds of L_p spaces but the distance being minimized in the geometric median's definition isn't L_1 but rather the L_2 (Euclidean) norm. What does the L_1 refer to?

Is it just a misnomer/an accident/a freak of nature/something historical?

estimation median point-estimation

Share Cite Improve this question Follow

edited Aug 28 '14 at 12:46 amoeba 89.8k 28 262 311

asked Aug 26 '14 at 3:01 Fixed Point 539 1 6 14

1 "L_1" really is the L_1 norm. The median is any point at which the sum of the L_1 norms between the points and the median is minimal, just as the mean is any point at which the sum of the L_2 norms and the mean ("centroid") is minimized. – whuber ♦ Aug 26 '14 at 22:50

1 @whuber The geometric median doesn't minimize the L_1 norm. It minimizes the L_2 norm. And the mean doesn't minimize the L_2 norm but rather the square of the L_2 norm. That is why the multivariate mean is also called the centroid. – Fixed Point Aug 26 '14 at 22:54

I said L_2 norm for the mean, right? And you're correct--I meant the square of it. (That nicety won't apply to the discussion of the L_1 norm, of course.) – whuber ♦ Aug 26 '14 at 22:55

Add a comment

2 Answers

Active Oldest Votes

6

Given a collection $\{\mathbf{x}^{(i)}\}$ of N points in $\mathbb{R}^m$, their geometric median is a point $\mathbf{y}$ minimizing the sum of Euclidean distances to each point:

$$L = \sum_{i=1}^{N} \|\mathbf{x}^{(i)} - \mathbf{y}\|.$$

Each Euclidean distance in this sum is indeed a L_2 norm of a vector in $\mathbb{R}^m$. But consider a vector $\mathbf{d} \in \mathbb{R}^N$, whose coordinates are given by these distances: $d_i = \|\mathbf{x}^{(i)} - \mathbf{y}\|$. Then the same cost function L can be equivalently written as the L_1 norm of this vector:

$$L = \sum_{i=1}^{N} d_i = \sum_{i=1}^{N} |d_i| = \|\mathbf{d}\|_1.$$

That is, I believe, why the geometric median is called a L_1 estimator.

Share Cite Improve this answer Follow

edited Feb 11 '15 at 14:51

answered Aug 28 '14 at 12:47 amoeba

amoeba
89.8k 28 262 311

Add a comment

0

[Response rewritten]

I think I was too confusing, I apologize for that. Now I am trying to give a proper answer.

We know that median minimize the L_1 norm. The formula is

$$L_1 = \underset{y \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} |x_i - y|$$

Also the mean minimize the L_2 norm. Again, the formula is

$$L_2 = \underset{y \in \mathbb{R}}{\arg\min} \sum_{i=1}^{n} (x_i - y)^2$$

In plain English we say that the median minimize the sum of distances and the mean minimize the sum of squares of those distances. We note also that we are in $\mathbb{R}$.

My idea is that because we are in $\mathbb{R}$, the distance function can be any particular case of *p-norm*, the result would be the same. So I generalize by saying that the distance is *p-norm* (it might be any type of distance in fact) and to finish quicker we move to $\mathbb{R}^m$ at the same time

$$L_1 = \underset{y \in \mathbb{R}^m}{\arg\min} \sum_{i=1}^{n} |(\sum_{j=1}^{m} (x_{i,j}^p - y_j^p))^{\frac{1}{p}}| = \underset{y \in \mathbb{R}^m}{\arg\min} \sum_{i=1}^{n} \|x_i - y\|_p$$

What is important here is that it does not matter what is the value for p, it will be an L_1. [Note, as suggested by @amoeba, there are two norms, one inside another; the first one is L_1 applied on distances, and a nested one applied on the elements of the vectors in $\mathbb{R}^m$].

Going back to your original question, the *geometric median* is defined as the point in Euclidean space which minimize the sum of distances. I believe the reason for the word geometric comes from Euclidean space and Euclidean distance (which is $\|.\|_2$) and minimize the sum of distances (not the squares as in the case of an L_2 estimator), so

$$GM = L_1 = \underset{y \in \mathbb{R}^m}{\arg\min} \sum_{i=1}^{n} \|x_i - y\|_2$$

As a final note we might choose to minimize:

- the sum of Manhattan distances (L_1 and $\|.\|_1$ as distance)
- the sum of squares of Manhattan distances (L_2 and $\|.\|_1$ as distance)
- the sum of Euclidean distances (geometric median)
- the sum of squares of Euclidean distances (L_2 and $\|.\|_2$ as distance) and so on.

Share Cite Improve this answer Follow

edited Aug 27 '14 at 22:58

answered Aug 26 '14 at 6:27

rapaio
6,190 24 44

1 First, all metrics by definition are non-negative so most of the absolute values are irrelevant. Second, this answer doesn't make any sense to me. Please see my addendum to the question for refutation. I am not the downvoter but I am very inclined to do so. – Fixed Point Aug 26 '14 at 22:49

(You got in here too fast for me, Fixed!) I downvoted this reply because it confuses the issue. When an L_p norm is of concern, as is the case for the median, a Euclidean (L_2) norm is irrelevant. Although there appears to be a correct answer buried here, it is hard to find because of the references to Euclidean space. – whuber ♦ Aug 26 '14 at 22:54

@whuber I don't agree with this comment. The norm is very relevant. In a dimension higher than one, the L_1 minimizer will be different than an L_2 minimizer which will be different than an L_3 minimizer ... which will be different than an L_∞ minimizer. – Fixed Point Aug 26 '14 at 23:06

1 @whuber Repeating, in $\mathbb{R}$, all p-norms are equal so the univariate median minimizes L_1 and L_2 and all other L_p norms. The geometric median, which is one generalization of the univariate median to higher dimensions, minimizes L_2 distance which is also called the Euclidean distance. The geometric median does NOT minimize L_1 distance. Take a look at wikipedia or any of your favorite papers/books talking about multivariate

Linked

38 Is there an accepted definition for the median of a sample on the plane, or higher ordered spaces?

Related

9 Expression for the median of a sum of half-Cauchy random variables

8 When to use sample median as an estimator for the median of a lognormal distribution?

6 Is the sample mean a better point estimate of the population median than the sample median?

21 Why is the James-Stein estimator called a "shrinkage" estimator?

0 Why does the formula for the median assume continous classes (grouped data)?

6 Proof that posterior median is the Bayes estimate of absolute loss?

L_1 distance. take a look at wikipedia or any of your favorite papers/books talking about multivariate medians. Why is there even an argument about that which is clear in the definition of the geometric median accepted by the community? – Fixed Point Aug 27 '14 at 2:46

1 Rapaio, I think there is one little piece that is missing from your answer even after the edit, and it makes it confusing for the @FixedPoint. There are *two* norms involved here. One norm defines the distance between points. If Euclidean distance is used (as in geometric median), then it is obviously a L_2 norm. But now consider distances from each $\mathbf{x}_{(i)}$ to $\mathbf{y}$ as one vector d_i, and *another* norm -- the norm of this vector $\mathbf{d}$. The geometric median minimizes the sum of its coordinates, i.e. its L_1 norm. I believe this answers the original question. – amoeba Aug 27 '14 at 21:57

Show **8** more comments

Your Answer

Not the answer you're looking for? Browse other questions tagged estimation median point-estimation or ask your own question.

Sign up or log in

Sign up using Google

Sign up using Facebook

Post as a guest

Name

Email

Required, but never shown