Subject: Re: Check for duplicate locations
Posted by Russell[1] on Mon, 22 Feb 2016 19:36:10 GMT
View Forum Message <> Reply to Message

On Thursday, February 18, 2016 at 11:48:38 AM UTC-5, Med Bennett wrote:
> I have X,Y,Z data for several thousand points that I need to check for duplicate locations. I cannot have duplicated locations in the sample data, as it breaks the kriging algorithm I am using. I've always used a brute force method of computing a distance function between each point and all subsequent points, and flagging any points for which the distance is zero, or some small threshold. This method is very slow for larger numbers of points, however. Does anyone have a method for doing this more efficiently? I've found simple methods for one-dimensional data, but not for points in 3-space.
>
> Thanks!

This is a tough answer to explain in just a few words....

Are the (x,y,z) values exactly the same? Either way, I would map them into a single coordinate (basically the inverse operation of array_indices) and compute the histogram. Consider a 2-d example...

(x,y)=(1,2)

and the maximum value of x,y could be (Nx,Ny)=(100,100). Then you can combine (x,y) into single value: xy = x+nx*y

Now use the histogram function on that new variable, and any bin in the histogram with more than a count of 2 has multiple entries. At that point you can do just about whatever to them.