Subject: Re: Need Some Advice on Seperating Out Some Data
Posted by btt on Tue, 08 Aug 2006 20:57:28 GMT
View Forum Message <> Reply to Message

rdellsy@gmail.com wrote:
> I considered that. Unfortunately, ambient conditions can vary the x and
> y positions of the data by as much as a factor of ten. That is why I am
> trying to figure out a method to compute it on the fly, since going
> through the process for just five movies can take up to half an hour,
> and dealing with fifty movies can be a full day's work.
> Thanks,
> Rob
>
> adisn123@yahoo.com wrote:
>> I used to have a similar problem. One of the simpliest thing that I did
>> was using a simple
>> linear equation such as y =ax + b.
>>
>> Overplot the linear eqaution in your original plot in such a way that
>> the linear line is placed
>> just above the red poligon (the data points that you want to throw out)
>> then
>>
>> simply you can throw out whatever the y values are below the linear
>> line.
>>
>>
>>
>>
>>
>> rdellsy@gmail.com wrote:
>>> http://photos1.blogger.com/blogger/4016/2263/320/graphroi.pn g
>>>
>>> The above is a plot of my data (minus the red polygon). I need to
>>> seperate the data inside the red polygon (real data) from the data
>>> outside the red polygon (noise, for lack of a better term) All of these
>>> points are already containted in an array. I'm just trying to figure
>>> out a way for the computer to automatically figure out what is noise
>>> and what isn't based on that plot distribution. Each data set is
>>> slightly different, but has the same overall distribution, and, for
>>> properly dialed in data, there is always that characteristic seperation
>>> between the good stuff and the bad stuff. Currently, we are manually
>>> setting x-boundaries and y-boundaries on our data.

Hi,

Just an end-of-the-day wildcard, but I would bin the data into a 2d
histogram (ala JD's HIST_ND or the built-in HIST_2D). Then I would try

to find the "saddle" between the data and noise. You'll have to fiddle with the binsize a bit to balance "lumping" and "splitting" - maybe that can be done dynamically. I dunno. But it should be quick.

It is an interesting problem that we have face here with flow cytometry - but we work the region manually as you do. I'll be interested to see what your final solution is.

Cheers,
Ben
