Subject: Re: Segregating data in bimodal distribution
Posted by Jeremy Bailin on Wed, 03 Aug 2011 15:37:36 GMT
View Forum Message <> Reply to Message

On 8/3/11 8:35 AM, Eric Hudson wrote:
> Hi,
>
> Is anyone aware of an IDL implemented algorithm for segregating data
> in a bimodal distribution into two groups?
>
> My data is such that I could do it manually (make a histogram, decide
> on a threshold between the two peaks in the histogram, then pull out
> the data above and below that into two separate groups). There isn't
> a true gap between the two peaks, but they are pretty well separated.
> The part which is non-obvious to me is to how to programmatically
> choose the threshold value. And since I have to do this on many data
> sets, where the threshold is going to be different for each, I prefer
> to not do it manually.
>
> Thanks,
> Eric
>
> PS In searching I found something called the KMM algorithm which
> seems like it would work, but I haven't found code for it.

Are the peaks well-represented by a known function (e.g. Gaussian)? If so, you could fit a bimodal Gaussian/whatever to the distribution and use the parameters of the fit to determine when the total is dominated by one or the other peak.

-Jeremy.