Subject: setting histogram bin sizes?
Posted by jeffnettles4870 on Tue, 22 Jul 2008 15:52:41 GMT
View Forum Message <> Reply to Message

Hi folks,

I'm looking for suggestions for a way to set bin sizes for a histogram when I don't know much about the data before calculating the histogram. Here's my situation: I'm putting together some code that takes a hyperspectral image cube and extracts a series of one-band parameters from the cube (band depth at a certain wavelength, etc.). In trying to assess which of these parameters is most useful for our particular application i thought about calculating a histogram for each parameter. The problem is that these parameter images (one band, floating point images per parameter) will not necessarily fall into the same range. Many have possible values of 0 - 1, but they won't necessarily take up that entire range. Some however, will not have possible values of 0 - 1, but could instead have numbers in the 10s or even hundreds. Some parameters have values that are actually in log space.

I know that I could simply set the NBINS keyword to HISTOGRAM(), but then the question would become how many bins to use? I did some quick searching, and there are a few attempts at calculating bin sizes or the number of bins on Wikipedia (http://en.wikipedia.org/wiki/Histogram). Short of any other information, i am going to use an equation from that page that is at least based on the standard deviation of the data. But, since I don't have a lot to go on, I would very much like to have input from anyone on this newsgroup who might have any suggestions for me.

Thanks,
Jeff

Subject: Re: setting histogram bin sizes?
Posted by dasergatskov on Sat, 26 Jul 2008 19:33:18 GMT
View Forum Message <> Reply to Message

On Jul 22, 10:52 am, "Jeff N." <jeffnettles4...@gmail.com> wrote:
> Hi folks,
>
> I'm looking for suggestions for a way to set bin sizes for a histogram
> when I don't know much about the data before calculating the
> histogram. Here's my situation: I'm putting together some code that
> takes a hyperspectral image cube and extracts a series of one-band
> parameters from the cube (band depth at a certain wavelength, etc.).
> In trying to assess which of these parameters is most useful for our

> particular application i thought about calculating a histogram for
> each parameter. The problem is that these parameter images (one band,
> floating point images per parameter) will not necessarily fall into
> the same range. Many have possible values of 0 - 1, but they won't
> necessarily take up that entire range. Some however, will not have
> possible values of 0 - 1, but could instead have numbers in the 10s or
> even hundreds. Some parameters have values that are actually in log
> space.
>
> I know that I could simply set the NBINS keyword to HISTOGRAM(), but
> then the question would become how many bins to use? I did some quick
> searching, and there are a few attempts at calculating bin sizes or
> the number of bins on Wikipedia (http://en.wikipedia.org/wiki/
> Histogram). Short of any other information, i am going to use an
> equation from that page that is at least based on the standard
> deviation of the data. But, since I don't have a lot to go on, I
> would very much like to have input from anyone on this newsgroup who
> might have any suggestions for me.
>
> Thanks,
> Jeff

I found Kevin Knuth (http://www.huginn.com/knuth/) paper
Optimal Data-Based Binning for Histograms
http://arxiv.org/abs/physics/0605197
to be quite useful.

Sincerely,

Dmitri.
--