
Subject: Re: Cluster analysis

Posted by [Chris.Jacobsen](#) on Fri, 03 Dec 2004 23:30:42 GMT

[View Forum Message](#) <> [Reply to Message](#)

hfrey@ssl.berkeley.edu (Harald) wrote in message

news:<51c9fb99.0412021550.41c2d53d@posting.google.com>...

- > The cluster analysis in IDL with clust_wts and cluster finds only 8
- > useful clusters. I can set n_clusters to a greater number, however all
- > cluster more than 8 useful ones are at locations, that do not contain
- > any of the samples. Is there any statistical reason that the number of
- > clusters can never exceed 8?
- >
- > I have a data set with 24 variables and 7680 samples and would expect
- > that I can easily find 20 or so different clusters.
- >
- > Does anybody have better IDL software that does the cluster analysis
- > properly?

In our work we started out using the "stock" IDL cluster routine but we have added to it a bit. Still, we have not changed the basic algorithm. We've found that preparation of the data can make a big difference. If the variation in variable X is 100 times bigger than the variation in variable Y, then the clustering (which looks at simple Euclidian distance) will not see the variation in Y very well. One approach is to subtract the mean of each variable, and apply a scale factor to the data in variable Y so that it is spread out over the same distance as in variable X.

With data preparation of that sort, the stock IDL routine can certainly find more than 8 clusters.

A paper on our work is at

http://xray1.physics.sunysb.edu/~micros/publications/papers/lerotic_ultramic_2004.pdf
