Subject: Robust covariance estimate
Posted by natha on Tue, 28 Mar 2017 19:19:50 GMT
View Forum Message <> Reply to Message

Hi all,

I am performing a Principal Component Analysis and I would like to use robust computation of the covariance matrix to avoid the outliers affect my results.
I've been reading a little bit about it and it seems that there are many approaches to compute robust covariance estimators.
I didn't find any code for that, only a a Python library which requires the input data matrix to be Gaussian distributed, which is not my case:
http://scikit-learn.org/stable/auto_examples/covariance/plot
_robust_vs_empirical_covariance.html

Any ideas or suggestions?
Thanks in advance for your help,
nata

Subject: Re: Robust covariance estimate
Posted by wlandsman on Tue, 28 Mar 2017 20:04:09 GMT
View Forum Message <> Reply to Message

I'm not much help except to say that the requirement for a Gaussian distribution is not very strict for most robust statistics programs (e.g. https://idlastro.gsfc.nasa.gov/ftp/pro/robust/). Presumably, you have a centrally concentrated distribution - otherwise how could one identify outliers? Assuming a Gaussian for the sole purpose of identifying outliers might be adequate. (If not Gaussian, do you know the true distribution of your data matrix?) -Wayne

On Tuesday, March 28, 2017 at 3:19:51 PM UTC-4, nata wrote:
> Hi all,
>
> I am performing a Principal Component Analysis and I would like to use robust computation of the covariance matrix to avoid the outliers affect my results.
> I've been reading a little bit about it and it seems that there are many approaches to compute robust covariance estimators.
> I didn't find any code for that, only a a Python library which requires the input data matrix to be Gaussian distributed, which is not my case:
> http://scikit-learn.org/stable/auto_examples/covariance/plot
_robust_vs_empirical_covariance.html
>
> Any ideas or suggestions?
> Thanks in advance for your help,
> nata