Subject: Re: Principal Componets Analysis
Posted by wlandsman on Wed, 05 Sep 2007 14:45:40 GMT
View Forum Message <> Reply to Message

On Sep 3, 8:33 pm, David Fanning <n...@dfanning.com> wrote:
>
> You can find the tutorial here:
>
> http://www.dfanning.com/code_tips/pca.html
>
> Any and all comments welcome.

Well, a minor historical comment about the different PCA conventions. The pcomp.pro procedure was introduced into IDL in 1996, but prior to that Immanuel Freedman had written a procedure pca.pro (http://idlastro.gsfc.nasa.gov/ftp/pro/math/pca.pro) based on a FORTRAN program by Fionn Murtagh.

When pcomp.pro was introduced, it took me a long time to prove that pca.pro and pcomp.pro gave the same results Below are the notes I wrote at the time:

****

The intrinsic IDL function PCOMP duplicates most most of the functionality of PCA, but uses different conventions and normalizations. Note the following:

(1) PCOMP requires a N_ATTRIB x N_OBJ input array; this is the transpose of what PCA expects
(2) PCA uses standardized variables for the correlation matrix: the input vectors are set to a mean of zero and variance of one and divided by sqrt(n); use the /STANDARIZE keyword to PCOMP for a direct comparison.
(3) PCA (unlike PCOMP) normalizes the eigenvectors by the square root of the eigenvalues.
(4) PCA returns cumulative percentages; the VARIANCES keyword of PCOMP returns the variance in each variable
(5) PCOMP divides the eigenvalues by (1/N_OBJ-1) when the covariance matrix is used.

****

And, yes, I verified that pca.pro also reproduces the results in your tutorial, but it requires even more adjustment than does pcomp.pro !