Computation of the Best-fitting Flat for a Set of Points

Gary D. Knott, Ph.D.
Civilized Software, Inc.
12109 Heritage Park Circle
Silver Spring MD 20906
Tel.: (301)-962-3711
email: csi@civilized.com
URL: www.civilized.com

In order to compute the best-fitting $k-1$ dimensional flat for a set of n points in k-space, we need only compute the eigenvector corresponding to the largest eigenvalue of the "covariance" matrix of the data points. This eigenvector is the unique normal direction which defines the desired flat. This flat is the hyperplane for which the sum of the squared perpendicular distances from the data points to the flat is minimal. The process of computing this flat is also known as principal components analysis, where we seek the subspace spanned by the last $k-1$ "principal component" eigenvectors. The eigenvector corresponding to the least eigenvalue defines the best-fitting line for the n data points; this line is known as the *principal axis* of the set of data points. Note that linear regression does not define the principal axis line of a set of data points; the linear regression line is that line that minimizes the sum of the squares of the vertical (y) distances from the data points to the line.

Below we show an example using MLAB to compute the best-fitting flat for the case $k = 2$. In this case there are only two orthogonal eigenvectors.

```
* q=read(data,50,2)
* n=nrows(q)
* qb=mean(q)
*
* q= q-(qb')^^n
*
* m=q'*q
* z=eigen(m) row (1,3,5)
* type z

   Z: a  3 by 2 matrix

  1: 681.733746    45.5137215
```

```
  2: .732144702    .681149128
  3: -.681149128   .732144702

*
* v=(z row 2)'
*
* fct g(x)=qb[2]+v[2]*(x-qb[1])/v[1]
* draw points(g,-1:11.6!2) color green lt (1,0,0,0,0,.0075,0)
* draw q lt none pt circle
* top title "best-fitting flat minimizing perpendicular error"
* view
* exit
```