Singular Value Decomposition.

Xoriginal Original data matrix

Center and scale columns so that sum of squared elements is 1, call this matrix Xnxk. Assume full rank, n>k.

If columns are values of random variables, X’X is correlation matrix.

1. Review

 Rkxk: matrix of k eigenvectors (the k columns)

D2: kxk diagonal matrix of eigenvalues

 Lnxn: matrix of n eigenvectors (the n columns)

Dn2: matrix of n eigenvalues (n-k are 0, others

match D2

Recall: 

1. Singular Value Decomposition of X (SVD)

SVD is defined as (you can always do this) so we have 

Singular values: diagonal elements of D = positive square roots of eigenvalues of . If we replace the smallest k-r of these with 0s then D becomes Dr and  is the best (least squares) rank r approximation to X.

So … we can sometimes replace the k columns of X with just r columns of L or LD and capture most of the variation in the X matrix. How many columns? (sum of largest r eigenvalues)/k is proportion of variability explained by the r columns – use as a

guide.

1. Principal components: Columns of matrix P where

 gives . Now  shows that each column of P is a linear combination of the X columns and shows that each column of P is a multiple (the corresponding singular value) of one of the orthogonal columns in L. If we replace the smallest k-r singular values in D with 0 then P contains only r nonzero columns and becomes Pr

The first r principal components

1. Are orthogonal to each other
2. Capture a proportion of the variation in X
3. Will give the same predictions in a regression of Y on P=Pk as in a regression of Y on X *if r=k* (i.e. if all are used).
4. Can be obtained in PROC PRINCOMP.

Regress Y on X. Coefficients are . Note that 

So … if r=k then we can completely recover b from bP from the principal component regression  and if r<k we can recover an approximation to b.