Multivariate and Functional Principal Components without Eigenanalysis

Jim Ramsay, McGill University
SSC 2012

PCA: The essential idea

- We have a N by n data matrix $\mathbf{X}$.
- We propose the reduced rank K bilinear model

$$\mathbf{X} = \mathbf{F}\mathbf{A}$$

  where
  - $\mathbf{A}$ is a K by n matrix of principal component coefficients, with $K << n$
  - $\mathbf{F}$ is a N by K matrix of principal component scores
- Usually $N >> n$, and the factor scores are interesting, but it's $\mathbf{A}$ that tells us what the core K components of variation are, to within a full rank linear transformation.
- The fundamental goal of PCA is to identify a linear subspace $\mathcal{R}^{\mathcal{K}}$.

What I'd like to do with PCA

- Provide GLM capability: PCA for mixtures of types of variables, using fitting criteria appropriate to each data type.
- Define a fitting strategy that recognizes PC scores **F** as nuisance parameters and PC components in **A** as structural parameters.
- Generalize PCA in many ways, but in this talk to implement partial least squares: an approximation of an external vector **y** via a K dimensional subspace $\mathcal{R}^{\mathcal{K}}$.

Why eigenanalysis gets in the way

- The singular value decomposition yields both **A** and **F**,
- but the usual procedure is to extract **A** from the eigenanalysis of $N^{-1}\mathbf{X}'\mathbf{X}$ or the correlation matrix **R**
- and then use regression analysis to obtain

$$\mathbf{F} = \mathbf{X}\mathbf{A}'(\mathbf{A}'\mathbf{A})^{-1}$$

- Eigenanalysis forces us to use least squares fitting for all variables.
- It treats the estimation of **F** and **A** symmetrically.
- It doesn't help for other data analysis problems such as partial least squares.

The parameter cascading strategy

- Parameter cascading defines nuisance parameters as *smooth* functions of structural parameters.
- We add smoothness to the least squares criterion for $\mathbf{F}$ given $\mathbf{A}$ by attaching penalty terms such as

$$J(\mathbf{F}|\mathbf{A},\mathbf{X}) = \|\mathbf{X} - \mathbf{FA})\|^2 + \lambda_1 \|\mathbf{F}'\mathbf{P}_1\mathbf{F}\|^2 + \lambda_2 \|\mathbf{FP}_2\mathbf{F}'\|^2.$$

- The minimizer $\hat{\mathbf{F}}(\mathbf{A})$ has a closed form expression.
- Order K matrix $\mathbf{P}_1$ and order N matrix $\mathbf{P}_2$ are often projectors onto complements of some pre-defined subspaces or special patterns.
- Smoothing parameters $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ allow us to control the emphasis that we place on the PC scores having these particular structures.

The fitting criterion for A

- This is defined in terms of only the PC coefficients **A**.
- For example, we could use

$$G(\mathbf{A}) = -\sum_{j}^{n} \ln L_j(\mathbf{A}|\mathbf{x}_j)$$

  where $-\ln L_j$ is the negative log likelihood appropriate to variable j and defined by data N-vector $\mathbf{x}_j$.
- Note that the gradient of G will depend on **A** both directly through its the partial derivative, and also via the N functions $\mathbf{f}_i(\mathbf{A})$.
- PCA is now defined as a nonlinear least squares problem with Kn parameters, as opposed to a bilinear least squares problem with $K(N+n)$ parameters.

The PCA/PLS hybrid criterion

- Keeping to LS fitting, we now use fitting criterion

$$G(\mathbf{A}|\mathbf{X}, \mathbf{y}) = (1-\gamma)\|\mathbf{X} - \mathbf{FA}\|^2 + \gamma\|\mathbf{y}'\mathbf{Q}(\mathbf{A})\mathbf{y}\|^2.$$

  where the relaxation parameter $\gamma \in [0, 1]$ and

$$\mathbf{Q}(\mathbf{A}) = \mathbf{I} - \mathbf{F}(\mathbf{A})[\mathbf{F}(\mathbf{A})'\mathbf{F}(\mathbf{A})]^{-1}\mathbf{F}(\mathbf{A})'.$$

- The second term measures the extent to which $\mathbf{y}$ is unpredictable from within the subspace defined by the PC loadings in $\mathbf{A}$.
- The boundary conditions $\gamma = 0$ and $\gamma = 1$ correspond to pure PCA and pure PLS, respectively.

Conclusions

- Parameter cascading not only regularizes the estimation of nuisance parameters in **F**,
- it re-defines PCA as a much lower dimensional fitting problem.
- Other variations of PCA are being investigated, including
  - Functional and hybrid multivariate/functional versions of PCA and PCA/PLS
  - Registration of functional data to their principal components