Electronic Journal of Statistics > Vol. 2 (2008) open journal systems

Functional principal components analysis via penalized rank one approximation

Jianhua Z. Huang, *Texas A&M Univeristy*
Haipeng Shen, *University of North Carolina at Chapel Hill*
Andreas Buja, *University of Pennsylvania*

Abstract

Two existing approaches to functional principal components analysis (FPCA) are due to Rice and Silverman (1991) and Silverman (1996), both based on maximizing variance but introducing penalization in different ways. In this article we propose an alternative approach to FPCA using penalized rank one approximation to the data matrix. Our contributions are four-fold: (1) by considering invariance under scale transformation of the measurements, the new formulation sheds light on how regularization should be performed for FPCA and suggests an efficient power algorithm for computation; (2) it naturally incorporates spline smoothing of discretized functional data; (3) the connection with smoothing splines also facilitates construction of cross-validation or generalized cross-validation criteria for smoothing parameter selection that allows efficient computation; (4) different smoothing parameters are permitted for different FPCs. The methodology is illustrated with a real data example and a simulation.

AMS 2000 subject classifications: Primary 62G08, 62H25; secondary 65F30.

Keywords: Functional data analysis, penalization, regularization, singular value decomposition.

Full Text: PDF

Huang, Jianhua Z., Shen, Haipeng, Buja, Andreas, Functional principal components analysis via penalized rank one approximation, Electronic Journal of Statistics, 2, (2008), 678-695 (electronic). DOI: 10.1214/08-EJS218.

References

Brown, L. D., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. (2005). Statistical analysis of a telephone call center: a queueing-science perspective. J. Amer. Statist. Assoc. 100 36–50. MR2166068

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. Numer. Math. 31 377–90. MR0516581

Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika 1 211–218.

Green, P. J. and Silverman, B. W. (1994). Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall. MR1270012

Hotelling (1933). Analysis of a complex of statistical variables in principal components. J. Educ. Psychol. 24 417–441, 498–520.

James G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data. Biometrika 87 587–602. MR1789811

Jolliffe, I. T. (2002). Principal Component Analysis. Springer-Verlag: New York, 2nd ed. MR2036084

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). Multivariate Analysis. Academic Press: London. MR0560319

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Phil. Mag. 6 1–7.

Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Springer-Verlag: New York, 2nd ed. MR2168993

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. J. Roy. Statist. Soc. Ser. B 53 233–243. MR1094283

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 57 253–59. MR1833314

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. Ann. Statist. 24 1–24. MR1389877

Shen, H. and Huang, J. Z. (2008). Interday forecasting and intraday updating of call center arrivals. Manufacturing and Service Operations Management 10 391–410.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100 577–90. MR2160561

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. Ann. Statist. 33 2873–903. MR2253106