Theoretical properties of Cook's PFC dimension reduction algorithm for linear regression

Oliver Johnson, *University of Bristol*

Abstract

We analyse the properties of the Principal Fitted Components (PFC) algorithm proposed by Cook. We derive theoretical properties of the resulting estimators, including sufficient conditions under which they are sqrt(n)-consistent, and explain some of the simulation results given in Cook's paper. We use techniques from random matrix theory and perturbation theory. We argue that, under Cook's model at least, the PFC algorithm should outperform the Principal Components algorithm.

AMS 2000 subject classifications: Primary 62H10; secondary 62E20.

Keywords: Principal Components, Principal Fitted Components, random matrix theory, regression.

Full Text: PDF

Johnson, Oliver, Theoretical properties of Cook's PFC dimension reduction algorithm for linear regression, Electronic Journal of Statistics, 2, (2008), 807-828 (electronic). DOI: 10.1214/08-EJS255.

References

[1] Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression. Statist. Sci. 22, 1, 1–26. MR2408655

[2] Critchley, F. (1985). Influence in principal components analysis. Biometrika 72, 3, 627–636. MR0817577

[3] Davidson, K. R. and Szarek, S. J. (2001). Local operator theory, random matrices and Banach spaces. In Handbook of the geometry of Banach spaces, Vol. I. North-Holland, Amsterdam, 317–366. MR1863696

[4] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philos. Trans. Roy. Soc. London Ser. A 222, 309–368.

[5] Fulton, W. (2000). Eigenvalues, invariant factors, highest weights, and Schubert calculus. Bull. Amer. Math. Soc. (N.S.) 37, 3, 209–249 (electronic). MR1754641

[6] Geman, S. (1980). A limit theorem for the norm of random matrices. Ann. Probab. 8, 2, 252–261. MR0566592

[7] Hotelling, H. (1933). Analysis of a complex statistical variable into principal components. J. Educational Psychology 24, 417–441.

[8] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist. 29, 2, 295–327. MR1863961

[9] Li, B., Zha, H., and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. Ann. Statist. 33, 4. MR2166556

[10] Li, K.-C. (1991). Sliced inverse regression for dimension reduction. J. Amer. Statist. Assoc. 86, 414, 316–342. With discussion and a rejoinder by the author. MR1137117

[11] Oman, S. D. (1991). Random calibration with many measurements: An application of stein estimation. Technometrics 33, 2, 187–195.

[12] Seber, G. A. F. (1984). Multivariate observations. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York. MR0746474

[13] Sibson, R. (1979). Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. J. Roy. Statist. Soc. Ser. B 41, 2, 217–229. MR0547248

[14] Weyl, H. (1912). Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen. Math. Ann. 71, 441–479. MR1511670

[15] Xia, Y., Tong, H., Li, W. K., and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 3, 363–410. MR1924297