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

A scale-based approach to finding effective dimensionality in manifold learning

Xiaohui Wang, *University of Virginia*
J. Steve Marron, *University of North Carolina at Chapel Hill*

Abstract

The discovering of low-dimensional manifolds in high-dimensional data is one of the main goals in manifold learning. We propose a new approach to identify the effective dimension (intrinsic dimension) of low-dimensional manifolds. The scale space viewpoint is the key to our approach enabling us to meet the challenge of noisy data. Our approach finds the effective dimensionality of the data over all scale without any prior knowledge. It has better performance compared with other methods especially in the presence of relatively large noise and is computationally efficient.

Keywords: Primary manifold learning, intrinsic dimension, scale space; secondary hypothesis test, multivariate analysis.

Full Text: PDF

Wang, Xiaohui, Marron, J. Steve, A scale-based approach to finding effective dimensionality in manifold learning, Electronic Journal of Statistics, 2, (2008), 127-148 (electronic). DOI: 10.1214/07-EJS137.

References

[1] Balasubramanian, M., Schwartz, E.L., (2002). The Isomap algorithm and topologi al stability. Science, 295, 7a.

[2] Becker, R., Chambers, J., and Wilks, A., (1988) The New S Language. Belmont, CA: Wadsworth.

[3] Bruske, J., Sommer, G. (1998). Intrinsic dimensionality estimation with optimally topology preserving maps. IEEE Trans. on PAMI, 20(5), 572–575.

[4] Camastra, F., Vinciarell, A., (2002). Estimating the intrisic dimension of data with a fractal-based approach. IEEE Trans. on PAMI, 24(10), 1404–1407.

[5] Chaudhuri, P., Marron, J.S., (1999). SiZer for exploration of structures in curves. Journal of the American Statistical Association, 94, 807–823. MR1723347

[6] Costa, J., Hero, A.O., (2004). Geodisic entropic graphs for dimension and entropy estimation in manifold learning. IEEE Trans. on Signal Processing, to appear. MR2085582

[7] Devroye, L., Györfi, L., Lugosi, G., (1996). A probabilistic theory of pattern recognition. Springer. MR1383093

[8] DeMers, D., and Cottrell, G., (1993). Nonliear dimensionality reduction. Advances in Neural Information Processing System, 5, 580–587.

[9] Donoho, D.L., Grimes, C., (2003). Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. PNAS, 100, no. 10, 5591-5596. MR1981019

[10] Fukunaga, K., Olsen, D.R., (1971). An algorithm for finding intrinsic dimensionality of data. IEEE Trans. on Computers, C-20, 176–183.

[11] Gmamadesolam. R., (1977). Methods for statistical data analysis of multivariate observations. John Wiley & Sons. MR0440802

[12] Hastie, T. (1984). Principal curves and surfaces. Technical report, Standord University, Dept. of Statistics.

[13] Hastie, T., Stuetzle. W., (1989). Principal curves. Journal of the American Statistical Association, 84, no. 406, 502–516. MR1010339

[14] Irie, b., and Kawato, M., (1990). Acquisition of internal representation by multi-layered perception. IEICE Trans. Inf. & Syst. (Japanese Edition), vol. J73-D-II, no. 8, 1173–1178.

[13] Jones, M.C., Marron, J.S., and Sheather, S.J., (1996). A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association, 91, no. 433, 401–407. MR1394097

[15] LeBlanc, M., Tibshirani, B., (1994). Adaptive principal surfaces. In Journal of the American Statisitical Association, 89, no. 425, 53–64.

[16] Levina, E., Bickel, P.J., (2005). Maximum likelihood estimation of intrisic dimension. In Advances in NIPS, 17, to appear.

[17] Lindeberg, T., (1993). Scale-space theory in computer vision. Kluwer Academic Publishers.

[18] Grassberger, P., Procaccia, I., (1983) Measuring the strangeness of strange attactors. Physica, D9, 189–208. MR0732572

[19] Ramsay, J.O., Silverman, S.W., (2002). Applied Functional Data Analysis, Springer, New York. MR1910407

[20] ter Haar Romeny, B.M., (2002). Front-end vision and multi-scale image analysis. Kluwer Academic Publishers.

[21] Roweis, S., Saul, L., (2000). Nonlinear dimensionality reduction by locally linear embedding. Science, 290, 2323-2326.

[21] Shepare, R.N., (1974). Representation of structure in similarity data: problems and prospects. Psychometrika, vol 39, No. 4, 373-421. MR0413412

[22] Schölkipf, B., Smola, A., and Müller, K., (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput., vol. 10, no. 5, 1299–1319.

[23] Smith, R.L., (1992). Optimal estimation of fractal dimension. Nonlinear Modeling and Forecasting, SFI in the Sciences of Complexity, Proc., Vol. XII, Eds. m. Casdagli & S. Eubank, Addison-Wesley, 115-135.

[24] Smith, R.L., (1992). Estimating dimension in noisy chaotic time series. Journal of the Royal Statistical Society Series B-Statistical Methodology, 54, 329-351. MR1160474

[25] Tenenbaum, J.B., de Silva, V., Langford, J.C., (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290, 2319–2322.

[26] Wang, H., Iyer, H., (2006). Application of local linear embedding to nonlinear exploratory latent structure. to apper in Psychometrika.

[27] Wang, X., (2004) A Scale-Based Approach to Finding Effective Dimensionality. Dissertation.

Electronic Journal of Statistics. ISSN: 1935-7524