

Zhu, Mu; Ghodsi, Ali

Automatic dimensionality selection from the scree plot via the use of profile likelihood.
(English) [Zbl 1157.62429](#)
Comput. Stat. Data Anal. 51, No. 2, 918-930 (2006).

Summary: Most dimension reduction techniques produce ordered coordinates so that only the first few coordinates need be considered in subsequent analyses. The choice of how many coordinates to use is often made with a visual heuristic, i.e., by making a scree plot and looking for a “big gap” or an “elbow.” In this article, we present a simple and automatic procedure to accomplish this goal by maximizing a simple profile likelihood function. We give a wide variety of both simulated and real examples.

MSC:

[62H25](#) Factor analysis and principal components; correspondence analysis
[62H99](#) Multivariate analysis

Cited in **15** Documents

Keywords:

[data compression](#); [denoising](#); [isomap](#); [latent semantic indexing](#); [manifold learning](#); [principal component analysis \(PCA\)](#); [resampling methods](#); [singular value decomposition \(SVD\)](#)

Software:

[ElemStatLearn](#); [TMG](#)

Full Text: [DOI](#)

References:

- [1] Cox, T.F.; Cox, M.A.A., Multidimensional scaling, (2001), Chapman & Hall New York · [Zbl 1004.91067](#)
- [2] Deerwester, S.; Dumais, S.T.; Landauer, T.K.; Furnas, G.W.; Harshman, R.A., Indexing by latent semantic analysis, *J. soc. inform. sci.*, 41, 6, 391-407, (1990)
- [3] Dumais, S.T., Improving the retrieval of information from external sources, *Behavior res. methods instrum. comput.*, 23, 2, 229-236, (1991)
- [4] Hastie, T.J.; Tibshirani, R.J.; Friedman, J.H., The elements of statistical learning: data-mining, inference and prediction, (2001), Springer Berlin · [Zbl 0973.62007](#)
- [5] Jolliffe, I.T., Principal component analysis, (2002), Springer Berlin · [Zbl 1011.62064](#)
- [6] Mardia, K.V.; Kent, J.T.; Bibby, J.M., Multivariate analysis, (1979), Academic Press New York · [Zbl 0432.62029](#)
- [7] McCullagh, P.; Nelder, J.A., Generalized linear models, (1989), Chapman & Hall New York · [Zbl 0744.62098](#)
- [8] Peng, F.; Schuurmans, D.; Wang, S., Augmenting naive Bayes classifiers with statistical language models, *Inform. retrieval*, 7, 3, 317-345, (2003)
- [9] Roweis, S.T.; Saul, L.K., Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290, 2323-2326, (2000)
- [10] Salton, G.; Buckley, C., Term weighting approaches in automatic text retrieval, *Inform. process. management*, 24, 5, 513-523, (1988)
- [11] Sprott, D.A., Statistical inference in science, (2000), Springer · [Zbl 0955.62006](#)
- [12] Tenenbaum, J.B.; de Silva, V.; Langford, J.C., A global geometric framework for nonlinear dimensionality reduction, *Science*, 290, 2319-2323, (2000)
- [13] Zeimpekis, D., Gallopoulos, E., 2004. TMG: a MATLAB toolbox for generating term-document matrices from text collections. Technical Report HPCLAB-SCG 1/6-04, Computer Engineering & Informatics Department, University of Patras, Greece.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.