SVD workshop abstracts

Scuola di Alta Formazione, Università degli Studi di Napoli "L'Orientale" Procida Island, Bay of Naples, 24-25 September 2015

Workshop presentation #1 Thursday 24 September, 15:00

Introduction to the SVD and its classic applications

Michael Greenacre, *Universitat Pompeu Fabra, Barcelona, Spain*

The singular value decomposition (SVD) is probably the most useful result in linear algebra, with many applications in all areas of multivariate statistical analysis. In this introduction I will

- define the SVD, reduced and complete, as well as its variants and generalizations,
- demonstrate some of its theoretical benefits,
- give several alternative geometric interpretations,
- briefly discuss its computation,
- show how its property of low rank matrix approximation is at the heart of a wide variety of multivariate methods.

References

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

Greenacre, M. (1984) *Theory and Applications of Correspondence Analysis*. Academic Press. Now available for free download from `www.carme-n.org`.

Puntanen, S., Styan, G. and Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models. Our Personal Top Twenty*. Springer.

Takane, Y. (2014) *Constrained Principal Component Analysis and Related Techniques*. Chapman & Hall / CRC.

Workshop presentation #2 Thursday 24 September, 16:00

Modern developments of the SVD

Trevor Hastie, ***Stanford University, Palo Alto, California, USA***

A low-rank SVD is one of those happy situations when an exact solution is available to a nonconvex optimization problem. Perturb the problem slightly, and you might be sunk. The nuclear norm is a convex relaxation of the rank of a matrix. In this talk I will discuss the use nuclear-norm regularization in the context of matrix completion, an important tool used with recommender systems.

In this talk I will discuss:

- missing values and the matrix completion problem,
- nuclear-norm regularization and the softImpute algorithm
- connections with alternating least squares
- large scale implementations of softImpute
- other applications of nuclear-norm regularization

References

Hastie, T. and Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity. The Lasso and Generalizations*. Chapman & Hall / CRC.

Hastie, T., Mazumder, R., Lee, J. and Zadeh, R. (2015) Matrix completion and low-rank SVD via fast alternating least squares. *Journal of Machine Learning Research (to appear)*

Workshop presentation #3 Friday 25 September, 9:30

SVD computation and updating for large data sets

Alfonso Iodice D'Enza, *University of Cassino, Italy*
Angelos Markos, *Democritus University of Thrace, Alexandropoulis, Greece*

For large-scale problems, computing the singular values and vectors of a matrix can be impractical or prohibitive. When data is produced at a high rate (*data flows*), classic SVD cannot be performed without retaining all data in memory. Numerous methods have been proposed to handle these problems in a computationally efficient way. In this introduction we will

- review incremental SVD computation techniques
- discuss areas of application
- focus on a family of sequential decomposition techniques with desirable properties
- describe incremental versions of representative multivariate methods
- present R implementations of these methods

References

Baker, C. G., Gallivan, K. A., & Van Dooren, P. (2012). Low-rank incremental methods for computing dominant singular subspaces. *Linear Algebra and its Applications*, *436*(8), 2866-2888.

Iodice D′ Enza, A., & Markos, A. (2015). Low-dimensional tracking of association structures in categorical data. *Statistics and Computing*, *25*(5), 1009-1022.

Workshop presentation #4 Friday 25 September, 11:30

Multiway extensions of the SVD

Pieter Kroonenberg, *University of Leiden, The Netherlands*

In this presentation I will discuss the extension of two-way SVD to three-way SVD variants – there is not a single one which has all the properties of the two-way SVD. I will mention multiway extensions in passing. The idea is to explain why it is necessary to consider such extensions at all, what the major characteristics are and which problems arise and which vistas open up. A sketch of technical issues involved will be given with appropriate references to the nitty-gritty of it all. An example will presented in some detail (a woman with a triple personality will be dissected) and available software will be exposed.

- Three-way data: profile data; three-mode rating data
- From two-way SVD to three-way SVD and higher-order (or multiway) SVD
- Why is two-way not enough? A three-way example.
- Useful graphics (joint biplots; single component line plots)
- Some more technical details (preprocessing, computation, missing data)
- Software (3WayPack, R programs, MatLab programs)

References

Books:

Smilde,A.K, Bro, R. & Geladi, P. (2005). *Multi-way Analysis: Applications in the Chemical Sciences. Chicester, UK: Wiley*

Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis. Hoboken, US: Wiley.*

Some important reviews:

Açar, E., & Yener, B. (2009). Unsupervised multiway data analysis: A literature survey. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 6–20.

Bro, R.(1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38,* 149–171

De Lathauwer, L. , De Moor, B. & Vandewalle, J. (2000). A Multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, *21*, 1253-1278.

Harshman, R. A., & Lundy, M. E. (1994). PARAFAC: Parallel factor analysis. *Computational Statistics & Data* Analysis, *18*, 39–72.

Kiers, H. A. L.; Van Mechelen, I. (2001) Three-way component analysis: Principles and illustrative application. *Psychological Methods*, *6*, 84-110.

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*, 455–500.