Numerical Linear Algebra for data-related applications

Yousef Saad[1]

[1]*Department of Computer Science and Engineering University of Minnesota Minneapolis, USA.*

Abstract

This tutorial will focus on linear algebra techniques for applications related to machine learning and other applications that rely of large data sets. A common tool that is exploited in solving data mining and machine learning problems is that of 'dimension reduction'. Dimension reduction is based on the precept that the observed data often lies in a noisy version of a low-dimensional subspace and so it is critical to work in this subspace not only to reduce computational cost but also to improve accuracy. The presentation will start with an overview of the key concepts and then illustrate dimension reduction methods with applications such as information retrieval, face recognition and matrix completion for recommender systems. One of the main difficulties in many of the methods based on dimension reduction is to find the inherent approximate rank of the data at hand. We will show how a few simple random sampling methods for computing spectral densities and counting eigenvalues can be used for this purpose. Among other topics, the tutorial will also cover applications of graph Laplaceans, such as clustering and image segmentation, as well methods for analyzing networks.