Principal component analysis for structured high dimensional data

Debashis Paul

Department of Statistics
University of California at Davis

Thursday, October 26, 2006, 4:00-5:00pm
Mathematics and Computer Science (MCS) Building, Room 149
111 Cummington Street, Boston

Tea and Cookies at 3:30pm in MCS 153

Abstract:
Increasingly we are confronting multivariate data with very high dimension and comparatively low sample size, e.g. in medical imaging, microarray analysis, speech and image recognition, atmospheric science, finance etc. In this talk we consider the problem of estimation of the principal components in situations where the dimension of the observation vectors are comparable to the sample size, even though the intrinsic dimensionality of the signal part of the data is small. It will be demonstrated that the standard principal component analysis can fail to provide good estimate of the eigenvectors of the population covariance matrix. However, if the eigenvectors corresponding to the bigger eigenvalues of the population covariance matrix are sparse in a suitable sense, then one can hope to get better estimates.A two-stage algorithm to efficiently deal with the problem of estimating the population eigenvectors will be proposed. The performance of the procedure will be demonstrated via some simulated and real data examples.

For directions and maps, please see *http://math.bu.edu/research/statistics/statseminar.html.*