A Compressed PCA Subspace Method for Anomaly Detection in High-dimensional Data

Eric D. Kolaczyk
Boston University
Department of Mathematics and Statistics
Boston, Massachusetts, USA
kolaczyk@bu.edu

Abstract

Random projection is a widely used method of dimension reduction. Its combination with standard techniques of regression and classification has been explored recently. Here we examine its use with subspace detection methods based on principal component analysis (PCA). Specifically, we show that, under appropriate conditions, with high probability the magnitude of the residuals of a PCA analysis of randomly projected data behaves nearly the same as the PCA residuals of the original data. Our results indicate the feasibility of applying subspace-based anomaly detection algorithms to randomly projected data, when the nature of the data covariance is effectively of sufficiently low dimension. We illustrate in the context of computer network traffic anomaly detection. This is joint work with Qi Ding.