COMPARISON OF DIMENSION REDUCTION METHODS USING RANDOM PROJECTIONS AND PRINCIPAL COMPONENT ANALYSIS

Abraham Engle, Cindy Luu, Ivvette Rios & Derrick Sund

Abstract

In a number of modern-day areas of research, notably including high-throughput microarray studies, an insufficient number of data points relative to the number of covariates is obtained to use traditional regression methods. In order to model the results obtained from such studies, dimension reduction techniques must be used in order to make existing algorithms more feasible; however, care must be taken in using these reductions, as they can significantly alter the structure of the data. The Johnson-Lindenstrauss lemma provides a guideline as to when projections that not alter the structure of the data can be made, but for many applications this is not adequate. In fact, the lemma may be somewhat crude, with significantly lower projections potentially possible. In this paper, we examine the dimension reduction method of Random Projections (RP), both in terms of how well it preserves the structure of high-dimensional data and in how well it performs when used with Cox Proportional Hazards (PH) Model in determining survival functions. Simulated data and corresponding coefficients, survival functions, and times were generated, with the predetermined functions compared with those generated by the regression. The results were compared with similar procedures used with the Principal Component Analysis (PCA) method of dimension reduction to evaluate their comparative effectiveness. Our results indicate that RP seems to offer no significant advantages over PCA, although their performance is strikingly similar. However, RP also fails to provide good preservation of the structure of the data, with significant changes in distances between data points.