The following are the abstracts for the RUSIS 2013

DIMENSION REDUCTION IN SURVIVAL ANALYSIS

ADRIAN CARBALLEIRA, KOURTNEY HOWELL, NOEL MARTINEZ
University of Arizona, Xavier University of Louisiana, University of Texas at El Paso

Abstract

With the onslaught of high-throughput and other data collection techniques, in recent years scientists have exponentially increased the amount of data that can be gathered from a given experiment. Consequently, this data, when expressed in an nxp matrix (representing p measurements taken from n sources), can be viewed mathematically as n vectors in R^p space, often with n << p. Due to the fact that p can often be quite large (thousands of dimensions), the "curse of dimensionality" along with computational restrictions can hinder attempts to extract meaningful information from the data. Here we compare two major dimension-reduction techniques through the use of Survival Analysis. One technique is Principal Components Analysis. The other is Random Projections. We use both of these techniques to project data into R^k where k < p. Dasgupta & Gupta (2002) and Achlioptas (2003) suggest lower bounds for k to be used with Random Projections. Here we use the statistical software R to determine how conservative these bounds are. We compare several different random projections in their abilities to reduce dimensionality of high-dimensional data. We also compute survival curves before and after dimension reduction for both PCA and Random Projections, calculating Bias and Mean Squared Error in order to compare the two.

Censored Data and Completion Methods

Evan Breikss, Aracely Alcala, Cristina Retamoza
Colorado School of Mines, St. Mary's University, University of Arizona

Abstract

In survival analysis the Kaplan-Meier and Nelson-Aalen estimators are utilized to include censored data. When comparing the operating characteristics presented by the Kaplan-Meier estimator to the empirical estimator, that is, in the absence of censored data, we determined that the Kaplan-Meier estimator performs the best when including censored data. In the case where the last data point of a study is censored, the completion methods proposed by B. Efron, R.D Gill, Brown et.al. and Z. Chen and E. Phadia attempt to provide information that may lie beyond the last given point. After running various distribution combinations and censoring rates we noticed that the Kaplan-Meier estimator had the lowest bias and the Nelson-Aalen estimator had the lowest mean square error. Overall, the tail completion proposed by Brown et al. appeared the most often with the lowest bias and mean square error. In comparing the tightness of two bias bounds, we noticed that the bound proposed by B. Efron is a tight lower bound while for the bound proposed by R.D Gill is a loose upper bound.