Penalized Fits to a Multi-way Layout with Multivariate Responses

Author: Rudolf Beran, University of California, Davis

Abstract:

Consider a complete k-way layout of d-dimensional mean vectors, in which each mean vector is an unknown function of k real-valued covariates whose values are known. The covariates may be either ordinal or nominal. There is at least one observation with error on each of the unknown mean vectors. The problem is to estimate the mean vectors efficiently, without making any assumption about the function that relates them to the known covariates. Both theory and practice have made it clear that the unconstrained least squares estimator of the mean vectors is unsatisfactory unless the data provides substantial replication.

This talk defines a candidate class of penalized least squares (PLS) estimators suitable for the problem. A separate quadratic penalty term is devised for each of the main effects and interactions in the MANOVA decomposition of the mean vectors. The construction of the penalty terms draws on vague notions about the unknown function that links the means to the covariates. Before being summed, each penalty term is weighted by right multiplication with a d x d symmetric positive semidefinite matrix. The candidate PLS estimators thereby accomplish, as special cases, both MANOVA submodel selection and dimensionality reduction.

The matrix penalty weights are chosen to minimize estimated quadratic risk over the candidate class of PLS estimators. It is shown that, as the number of cells in the complete k-way layout tends to infinity, the candidate PLS estimator with smallest estimated risk converges, in loss or risk, to the candidate estimator with smallest actual loss or risk. The asymptotics make no assumptions about the unknown d-dimensional mean vectors in the k-way layout and require no replication. A case study on multivariate response data illustrates how the proposed adaptive estimator works.