

The abstract:

An important step in developing individualized treatment strategies is to correctly identify subgroups of a heterogeneous population, so that different treatments can be given to different subgroups accordingly. In this paper, we consider the situation with samples drawn from a population consisting of subgroups with different means, along with certain covariables. We propose a penalized approach for subgroup analysis based on a regression model, in which heterogeneity can be represented by using an individual intercept value for each observation. We apply concave penalty functions to pairwise differences of the intercepts, which automatically lead to partitioning the observations into subgroups. We develop an alternating direction method of multipliers algorithm with concave penalties to implement the proposed approach and demonstrate its convergence. We also establish the theoretical properties of our proposed estimator and determine the order requirements of the minimal difference of signals between groups in order to recover them. These results provide a sound basis for making statistical inference in subgroup analysis. Our proposed method is further illustrated by simulation studies and the Cleveland heart disease dataset.