Building a Robust Explanatory Model from a Large Data Set

David B. Kim

Department of Mathematical Sciences
United States Military Academy
West Point, New York

ARL Sponsor: Barry A. Bodt

U.S. Army Research Laboratory
Aberdeen Proving Ground, Maryland

ABSTRACT:

Explosive growth in the size of modern data sets is found not only in the number of records but also in the number of variables each record carries. Building any kind of predictive or explanatory model from such a data set presents a novel challenge. We present a method of building an explanatory model based on David Scott's parametric L2E criterion. The empirical evidence suggests that the method exhibits a feature which enables presence of outliers to be easily detected and can also be used to select explanatory variables with (very much counter-intuitively) less ambiguity when the outliers are present than otherwise. We apply the method to develop plan evaluation metrics through analyses of simulation runs of an urban combat scenario.

KEY WORDS: Robust statistics, Minimum distance estimation, Variable selection, OneSAF

CONTACT: David B. Kim, USMA, West Point, NY 10996
Tel: (845) 938-4018 email: ad1384@exmail.usma.army.mil

Barry A. Bodt, Ph.D., ARL, APG, MD, 21005-5067
Tel: (410) 278-6659 email: babodt@arl.army.mil