

## STATISTICAL ANALYSIS AND DATA MINING

Original Article

**Iterative selection using orthogonal regression techniques**Bradley Turnbull, Subhashis Ghosal , Hao Helen Zhang

First published: 11 December 2013

<https://doi.org/10.1002/sam.11212>

About



Access

als



ire

## Abstract

High dimensional data are nowadays encountered in various branches of science. Variable selection techniques play a key role in analyzing high dimensional data. Generally two approaches for variable selection in the high dimensional data setting are considered—forward selection methods and penalization methods. In the former, variables are introduced in the model one at a time depending on their ability to explain variation and the procedure is terminated at some stage following some stopping rule. In penalization techniques such as the least absolute selection and shrinkage operator (LASSO), an optimization procedure is carried out with an added carefully chosen penalty function, so that the solutions have a sparse structure. Recently, the idea of penalized forward selection has been introduced. The motivation comes from the fact that the penalization techniques like the LASSO give rise to closed form expressions when used in one dimension, just like the least squares estimator. Hence one can repeat such a procedure in a forward selection setting until it converges. The resulting procedure selects sparser models than comparable methods without compromising on predictive power. However, when the regressor is high dimensional, it is typical that many predictors are highly correlated. We show that in such situations, it is possible to improve stability and computational efficiency of the procedure further by introducing an orthogonalization step. At each selection step, variables potentially available to be selected in the model are screened on the basis of their correlation with variables already in the model, thus preventing unnecessary duplication. The new strategy, called the Selection Technique in Orthogonalized Regression Models (STORM), turns out to be extremely successful in reducing the model dimension further and also leads to improved predicting power. We also consider an aggressive version of the STORM, where a potential predictor will be permanently removed from further consideration if its regression coefficient is estimated as zero at any stage. We shall carry out a detailed simulation study to compare the newly proposed method with existing ones and analyze a real dataset. © 2013 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining*, 2013

**About Wiley Online Library**



**Help & Support**



**Opportunities**



**Connect with Wiley**



Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

**WILEY**