

## STATISTICAL ANALYSIS AND DATA MINING

Original Article

**Penalized regression and risk prediction in genome - wide association studies**Erin Austin, Wei Pan , Xiaotong Shen

First published: 22 February 2013

<https://doi.org/10.1002/sam.11183>

Cited by: 7



About



Access

JLS



ire

## Abstract

An important task in personalized medicine is to predict disease risk based on a person's genome, e.g. on a large number of single - nucleotide polymorphisms (SNPs). Genome - wide association studies (GWAS) make SNP and phenotype data available to researchers. A critical question for researchers is how to best predict disease risk. Penalized regression equipped with variable selection, such as least absolute shrinkage and selection operator (LASSO) and smoothly clipped absolute deviation (SCAD), is deemed to be promising in this setting. However, the sparsity assumption taken by the LASSO, SCAD, and many other penalized regression techniques may not be applicable here: it is now hypothesized that many common diseases are associated with many SNPs with small to moderate effects. In this article, we use the GWAS data from the Wellcome Trust Case Control Consortium (WTCCC) to investigate the performance of various unpenalized and penalized regression approaches under true sparse or non - sparse models. We find that in general penalized regression outperformed unpenalized regression; SCAD, truncated  $L_1$ -penalty (TLP), and LASSO performed best for sparse models, while elastic net regression was the winner, followed by ridge, TLP, and LASSO, for non - sparse models. © 2013 Wiley Periodicals, Inc. Statistical Analysis and Data Mining, 2013

## Citing Literature

About Wiley Online Library Help & Support Opportunities 

