## **REGULARIZED GENERAL LINEAR REGRESSION FOR GENOME WIDE ASSOCIATION ANALYSES**

## S. Wagenpfeil<sup>1</sup>

<sup>1</sup>Institute for Medical Statistics and Epidemiology, TU Munich, Germany

Email: stefan.wagenpfeil@imse.med.tu-munechen.de

Regularized regression can be considered as generalization of classical regression approaches via penalization. This modification allows for problems where the number of regression parameters and independent variables p to be estimated is much larger than the number of individuals or independent measurements n at hand, i.e p > n. In genetics, for example, the problem arises when analysing genome wide association studies with today p = 500.000 the number of single-nucleotide polymorphisms and n the number of individuals in a study with usually n < 1.000, thus p >> n. We introduce a computational and numerical efficient way of inference for regularized general linear regression models giving an explicit analytic solution for the inversion problem coming from regression parameter estimation that can deal with extremely large-scaled covariate data. First simulation results show that computation time can tremendously be reduced from (extrapolated) hundreds or thousands of years to only minutes. The explicite solution has further essential numerical advantages concerning round-off errors. On the whole the approach offers a way of analysing genome wide association data in a general and unified statistical modelling framework.