Directed Acyclic Graphs and the use of Linear Mixed models

S.H. Heisterkamp [1,2]

1 Global Biometrics, Gruenenthal, Aachen, Germany

2 Groningen Bioinformatics Centre, University of Groningen, The Netherlands

The construction of Bayesian Networks and especially Directed Acyclic Graphs (DAG in genomics is a versatile tool to unravel the mechanism of action of combined genes on e.g. drug response. Presently most algorithms are designed to find DAG's from a large number of genetic markers, be it expression of genes in microarrays, SNP's or other relevant genetic information. As with other brute force tools as multiple regression, ridge regression , lasso and elastic net application of these methods have a high risk of over-fitting and chance findings. In our view the main reason for that is not the problem of the technique perse, but its lack of successfully including biological relevant information.

We propose to use prior information on candidate DAG's for certain supposed mechanisms of action using do-calculus to find causal pathways. Because a full Bayesian approach using MCMC is rather costly and complex to program, we propose an empirical Bayes solution. In this approach we will exploit the connection between DAG, mixed models and penalized likelihood and use both L_2 (ridge) and L_1 (lasso) and L_0 (count) penalties in the context of a mixed model.

To fit these models, methods are added onto the standard mixed-model methods of S-Plus and R.

References

Hoerl, A., Kennard, R. (1988) Ridge Regression. In *Encyclopedia of Statistical Sciences*, vol. 8, 129-136. NewYork: Wiley.

Kjaerulff U.B., Madsen A.L. *Bayesian Networks and Influence Diagrams*, (2008). New York: Springer Verlag.

Pearl, J. Causality: *Models Reasoning and Inference* (2000). Cambridge, Cambridge University Press

Pinheiro, J. C., Bates, D.M. (2001) *Mixed Models in S and S-Plus,* 2de print, Springer Verlag, NY.

Tibshirani, R. (1995) Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc. B* , **58** , 267-288.

Zou, H., Hastie, T. (2005) Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, **67**, Part 2, 301-320.