Computationally tractable fitting of graphical models:
the cost and benefits of decomposable Bayesian and penalized likelihood approaches.

A thesis presented in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Albany,

New Zealand.

Anne Marie Fitch

2012

Abstract

Gaussian graphical models are a useful tool for eliciting information about relationships in data with a multivariate normal distribution. In the first part of this thesis we demonstrate that partial correlation graphs facilitate different and better insight into high-dimensional data than sample correlations. This raises the question of which method one should use to model and estimate the parameters. In the second, and major part, we take a more theoretical focus examining the costs and benefits of two popular approaches to model selection and parameter estimation (penalized likelihood and decomposable Bayesian) when the true graph is non-decomposable.

We first consider the effect a restriction to decomposable models has on the estimation of both the inverse covariance matrix and the covariance matrix. Using the variance as a measure of variability we compare non-decomposable and decomposable models. Here we find that, if the true model is non-decomposable, the variance of estimates is demonstrably larger when a decomposable model is used. Although the cost in terms of accuracy is fairly small when estimating the inverse covariance matrix, this is not the case when estimation of the covariance matrix is the goal. In this case using a decomposable model caused up to 200-fold increases in the variance of estimates.

Finally we compare the latest decomposable Bayesian method (the feature-inclusion stochastic search) with penalized likelihood methods (graphical lasso and adaptive graphical lasso) on measures of model selection and prediction performance. Here we find that graphical lasso is clearly outclassed on all measures by both adaptive graphical lasso and feature-inclusion stochastic search. The sample size and the ultimate goal of the estimation will determine whether adaptive graphical lasso or feature-inclusion stochastic search is better.

Acknowledgements

First and foremost thanks to my supervisor Beatrix Jones . I have greatly benefited from her encouragement and help along the way. Without her I would never have started let alone finished! Thanks also to my co-supervisor James Curran, especially for his help with R-graphics and LaTeXcode.

Three authors made their code available to us. Thanks to Xianghong Zhou for the shortest path analysis code used by the authors of Zhou *et al.* (2002); James Scott for his C++ code for feature-inclusion stochastic search and Yang Feng for his R-code for adaptive graphical lasso.

Thanks to Robert McKibbin and Tony Norris (Heads of Institute) and Jeff Hunter, Howard Edwards and Marti Anderson (Statistics discipline leaders) who in these roles have both facilitated the opportunity for me to combine study and teaching and also gave much encouragement along the way. Thanks to fellow student Insha Ullah for sharing his R-code for finding a perfect elimination ordering. Thanks also to the rest of the Massey Statistics and Mathematics teams for your encouragement, especially over the write-up months.

Last, but certainly not least, a big thank you to my family. Thanks to Harmen for his help in getting Figure 3.3 into a printable form. Thanks to Nigel, Miriam and Harmen, Joanna, and Andrew, without their support and encouragement I could not have done this.

Whaia te iti kahurangi ki te tuohu koe me he maunga teitei.

Contents

Appendices

List of Figures

List of Tables