VARIATIONAL ALGORITHMS FOR APPROXIMATE BAYESIAN INFERENCE

by

Matthew J. Beal

M.A., M.Sci., Physics, University of Cambridge, UK (1998)

The Gatsby Computational Neuroscience Unit
University College London
17 Queen Square
London WC1N 3AR

A Thesis submitted for the degree of
Doctor of Philosophy of the University of London

May 2003

Abstract

The Bayesian framework for machine learning allows for the incorporation of prior knowledge in a coherent way, avoids overfitting problems, and provides a principled basis for selecting between alternative models. Unfortunately the computations required are usually intractable. This thesis presents a unified variational Bayesian (VB) framework which approximates these computations in models with latent variables using a lower bound on the marginal likelihood.

Chapter 1 presents background material on Bayesian inference, graphical models, and propagation algorithms. Chapter 2 forms the theoretical core of the thesis, generalising the expectation-maximisation (EM) algorithm for learning maximum likelihood parameters to the VB EM algorithm which integrates over model parameters. The algorithm is then specialised to the large family of conjugate-exponential (CE) graphical models, and several theorems are presented to pave the road for automated VB derivation procedures in both directed and undirected graphs (Bayesian and Markov networks, respectively).

Chapters 3-5 derive and apply the VB EM algorithm to three commonly-used and important models: mixtures of factor analysers, linear dynamical systems, and hidden Markov models. It is shown how model selection tasks such as determining the dimensionality, cardinality, or number of variables are possible using VB approximations. Also explored are methods for combining sampling procedures with variational approximations, to estimate the tightness of VB bounds and to obtain more effective sampling algorithms. Chapter 6 applies VB learning to a long-standing problem of scoring discrete-variable directed acyclic graphs, and compares the performance to annealed importance sampling amongst other methods. Throughout, the VB approximation is compared to other methods including sampling, Cheeseman-Stutz, and asymptotic approximations such as BIC. The thesis concludes with a discussion of evolving directions for model selection including infinite models and alternative approximations to the marginal likelihood.

Acknowledgements

I am very grateful to my advisor Zoubin Ghahramani for his guidance in this work, bringing energy and thoughtful insight into every one of our discussions. I would also like to thank other senior Gatsby Unit members including Hagai Attias, Phil Dawid, Peter Dayan, Geoff Hinton, Carl Rasmussen and Sam Roweis, for numerous discussions and inspirational comments.

My research has been punctuated by two internships at Microsoft Research in Cambridge and in Redmond. Whilst this thesis does not contain research carried out in these labs, I would like to thank colleagues there for interesting and often seductive discussion, including Christopher Bishop, Andrew Blake, David Heckerman, Nebojsa Jojic and Neil Lawrence.

Amongst many others I would like to thank especially the following people for their support and useful comments: Andrew Brown, Nando de Freitas, Oliver Downs, Alex Gray, Yoel Haitovsky, Sham Kakade, Alex Korenberg, David MacKay, James Miskin, Quaid Morris, Iain Murray, Radford Neal, Simon Osindero, Lawrence Saul, Matthias Seeger, Amos Storkey, Yee-Whye Teh, Eric Tuttle, Naonori Ueda, John Winn, Chris Williams, and Angela Yu.

I should thank my friends, in particular Paola Atkinson, Tania Lillywhite, Amanda Parmar, James Tinworth and Mark West for providing me with various combinations of shelter, companionship and retreat during my time in London. Last, but by no means least I would like to thank my family for their love and nurture in all my years, and especially my dear fiancée Cassandre Creswell for her love, encouragement and endless patience with me.

The work in this thesis was carried out at the Gatsby Computational Neuroscience Unit which is funded by the Gatsby Charitable Foundation. I am grateful to the Institute of Physics, the NIPS foundation, the UCL graduate school and Microsoft Research for generous travel grants.

Contents

List of figures

List of tables

List of Algorithms