Bayesian Learning for Neural Networks is no longer available on-line. A revised version with some new material has now been published by Springer-Verlag New York. You should now be able to order a copy at 1-800-Springer or orders@springer-ny.com. (However, since the book was just published (23 July 1996) it's possible that you could be a bit too early.) The ISBN is 0-387-94724-8. The price is $32.95 (US).

On the next page is the abstract of the original thesis.

Bayesian Learning for Neural Networks

Radford M. Neal

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy,
Graduate Department of Computer Science,
in the University of Toronto
Convocation of March 1995

Abstract

Two features distinguish the Bayesian approach to learning models from data. First, beliefs derived from background knowledge are used to select a prior probability distribution for the model parameters. Second, predictions of future observations are made by integrating the model's predictions with respect to the posterior parameter distribution obtained by updating this prior to take account of the data. For neural network models, both these aspects present difficulties — the prior over network parameters has no obvious relation to our prior knowledge, and integration over the posterior is computationally very demanding.

I address the first problem by defining classes of prior distributions for network parameters that reach sensible limits as the size of the network goes to infinity. In this limit, the properties of these priors can be elucidated. Some priors converge to Gaussian processes, in which functions computed by the network may be smooth, Brownian, or fractionally Brownian. Other priors converge to non-Gaussian stable processes. Interesting effects are obtained by combining priors of both sorts in networks with more than one hidden layer.

The problem of integrating over the posterior can be solved using Markov chain Monte Carlo methods. I demonstrate that the hybrid Monte Carlo algorithm, which is based on dynamical simulation, is superior to methods based on simple random walks.

I use a hybrid Monte Carlo implementation to test the performance of Bayesian neural network models on several synthetic and real data sets. Good results are obtained on small data sets when large networks are used in conjunction with priors designed to reach limits as network size increases, confirming that with Bayesian learning one need not restrict the complexity of the network based on the size of the data set. A Bayesian approach is also found to be effective in automatically determining the relevance of inputs.