Stochastic Neural Networks

Keith Rennolls, Alan Soper, Phil Robbins, Ray Guthrie.

CASSM, University of Greenwich, London SE18 6PF.

ABSTRACT

The two layer feed-forward neural net (the perceptron) may be regarded as a generalization of a multiple logistic regression model with multivariate output, if the threshold function is interpreted as a generalised linear model link. The multi-layer perceptron can be, and has been, used as a highly structured and parameterised non-linear model of expectation behaviour. The usual back propagation rules for training a neural net are equivalent to fairly simple methods for minimizing the error sum of squares at the outputs, ie. OLS. There has been little work done which uses the multi-layer perceptron as a statistical trend model in which the error distributions at the output nodes are statistically specified. For such non-linear models, OLS will be inefficient as well as biased. However, an appropriate specification of the stochastic error structure will lead to either weighted least squares or maximum likelihood estimates, which may be achieved either directly or by iteratively re-weighted least squares.

This general statistical approach allows the precision of trained weights to be estimated, as well as the correlation between these weights in the variance-covariance matrix. Such a feature is useful in examining the conjecture that different layers of the MLP have characterised different aspects of the training data set. Statistical methods can then also be used to prune the topology of the network using a step-wise approach.

Availability of the variance covariance matrix of the fitted weights allows the use of the trained MLP to make predictions. With simulated weights from the estimated distribution for the weight estimates prediction distributions may be obtained.

Steepest descent, and hence the back propagation algorithm, is one of the least efficient of optimisation algorithms available. The Fletcher Powell method is much more efficient for well behaved objective functions, whilst the Nelder-Mead algorithm is much faster, and more robust, for multi-modal functions. The use of simulated annealing is indicated for very badly behaved functions.

All of the above statistical features have been built in to a package which also implements the alternative training algorithms. A menu-based inter-face has also been developed for ease of usage. All features will be illustrated in the paper and in demonstration.