FLAT MINIMA

NEURAL COMPUTATION 9(1):1-42 (1997)

Sepp Hochreiter, TUM
Jürgen Schmidhuber, IDSIA

http://www.idsia.ch/~juergen

March 1996

Abstract:

We present a new algorithm for finding low complexity neural networks with high generalization capability. The algorithm searches for a ``flat'' minimum of the error function. A flat minimum is a large connected region in weight-space where the error remains approximately constant. An MDL-based, Bayesian argument suggests that flat minima correspond to ``simple'' networks and low expected overfitting. The argument is based on a Gibbs algorithm variant and a novel way of splitting generalization error into underfitting and overfitting error. Unlike many previous approaches, ours does not require Gaussian assumptions and does not depend on a ``good'' weight prior - instead we have a prior over input/output functions, thus taking into account net architecture and training set. Although our algorithm requires the computation of second order derivatives, it has backprop's order of complexity. Automatically, it effectively prunes units, weights, and input lines. Various experiments with feedforward and recurrent nets are described. In an application to stock market prediction, flat minimum search outperforms (1) conventional backprop, (2) weight decay, (3) ``optimal brain surgeon'' / ``optimal brain damage''. We also provide pseudo code of the algorithm (omitted from the NC-version).

Next: BASIC IDEAS / OUTLINE
Juergen Schmidhuber 2003-02-13

Back to Financial Forecasting page