Final Report on EPSRC Research Grant GR/L52093
Learning Fixed Example Sets in Multilayer Neural Networks

Summary

David Saad
Aston University, Birmingham B4 7ET

A general framework for analysing online learning in multilayer neural networks was developed several years ago by the principal investigator and collaborators, and has been very successful in obtaining an exact description of the learning dynamics in the limit of very large (infinite) example sets with respect to the number of free parameters (network weights), where training examples are sampled without repetition. However, this simple framework cannot be used for analysing learning characteristics which show up only in scenarios where the number of examples scales with the number of weights. Some of these effects are of great practical importance, especially in the case of noisy training data. The main difference between the two analyses is that, in the case of infinite training sets, no correlations are building up between the weights and the new examples, what allows one to carry out averages over each time step separately; on the other hand, when sampled with repetition, the examples are correlated with the evolving weights what makes the calculation extremely difficult.

The main aim of this project was to develop the theoretical framework for on-line learning from fixed example sets and to employ the new analytical description for investigating the effect of noisy training sets on overfitting. Secondary aims included the investigation of heuristic techniques currently in use, e.g., regularisation and early stopping, and a comparison between the theoretical results and those obtained in real-world tasks. The main achievements are summarised below.

Setting up the general framework: In the general scenario examined, a nonlinear model network (student) is trained on examples generated by an underlying unknown rule (teacher) represented by a similar model. We formulated the training problem within the Dynamical Replica Theory approach. This framework, which is completely general, enables one to analyse both continuous and discrete mappings and *any* learning rule as long as it is represented by some function (depending on the system parameters and the current example) multiplied by the pattern presented.

Extending the framework to the case of multilayer networks: While extending the main part of the analysis to the multilayer case is straightforward, the computation, which was just possible in the case of single layer networks is clearly infeasible in the case of multilayer networks. We therefore resort to the large α approximation, where α represents the ratio between the number of examples and the system size; the latter is particularly suitable to the multilayer case, since the main features of learning in multilayer networks, such as the breaking of internal symmetries and the asymptotic convergence, can be observed at sensible time scales only for relatively high α values. This framework was employed in studying both realisable and unrealisable scenarios where the model system has less hidden nodes than the teacher system.

Learning noisy data sets: We employed the same framework for studying learning from data sets corrupted by additive Gaussian output noise, assessing its effect on the learning dynamics and the evolution of both training and generalisation errors, demonstrating the advantages of early stopping. We also examined the effect of Gaussian regularizers in this scenario, showing the improvement obtained in network performance.

The project resulted in 13 publications, most of which are available through this web-page.

Acknowledgement: This research was supported by the EPSRC grant GR/L52093.