Capacity of Two-Layer Networks with Binary Weights

Chuanyi Ji* and Demetri Psaltis**
*Department of Electrical Computer and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
**Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

Abstract

The capacity C_b of two layer ($N-2L-1$) feedforward neural networks is shown to satisfy the relation $O(\frac{W}{lnW}) \leq C \leq O(W)$. Here $N-2L-1$ stands for the networks with N input units, $2L$ hidden units and one output unit. W is the total number of weights of the networks. The weights take only binary values and the hidden units have integer thresholds.

Summary

The motivation for this work comes from hardware implementation of neural networks. When weights of neural networks are implemented, both their accuracy and magnitude have to be limited. Then a natural question to ask is whether the learning capability of neural networks will thus be affected.

Learning capability of neural networks can be characterized by their information capacity[2], which is defined as the total number of dichotomies implementable by a class of networks of the same architecture. The capacity C of two layer $N-L-1$ feedforward networks with *analog* weights has been shown to satisfy the relation $O(W) \leq C \leq O(WlnL)$[1]. Here W, L and N are the total number of weights, the number of hidden units and the input dimension, respectively. It remains an open question, however, what the capacity of multilayer networks would be if their weights can only take discrete values. In this work we answer this question by evaluating the capacity of two layer $N-2L-1$ feedforward networks (N inputs, $2L$ hidden units and 1 output) with binary weights and integer thresholds for the hidden units.

Specifically, upper and lower bounds for the capacity C_b of such networks are established in two steps. First, the statistical capacity[3] of a specifically constructed network is evaluated and found to be $O(\frac{W}{lnW})$, where W is the total number of weights of the network. It is used as a lower bound for the capacity C_b. Then an upper bound is obtained through a simple counting argument, and shown to be $O(W)$. Therefore, we have $O(\frac{W}{lnW}) \leq C_b \leq O(W)$.

This result shows that reducing the analog weights to only binary values, the capacity of two-layer networks is reduced by at most a log factor. This is consistent to what has been found for a single neuron with binary weights[4]. Therefore, even with binary weights only, multilayer neural networks still have strong learning capability.

References

[1] E. Baum, " On the Capacity of Multilayer Perceptron," *J. of Complexity*, 1988.

[2] T.M. Cover, " Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition," *IEEE Trans. Elec. Comp.*, EC-14, pp 326-334, June, 1965.

[3] R.J . McEliece, E.C . Posner, E.R . Rodemich, S.S . Venkatesh, "The Capacity of the Hopfield Associative Memory," *IEEE Trans. Inform. Theory*, Vol. IT-33, No. 4, pp 461-482, July 1987.

[4] S. Venkatesh, "Directed Drift: A New Linear Threshold Algorithm for Learning Binary Weights On-Line," *Journal of Computer and Systems Sciences*, in press.