A Linear Model of a Learning Scheme

W.D. Wallis[1] and Chia-Lun J. Hu

Department of Mathematics and
Department of Electrical Engineering
Southern Illinois University
Carbondale, IL 62901-4408

Abstract. We discuss the learning problem in a two-layer neural network. The problem is reduced to a system of linear inequalities, and the solvability of the system is discussed.

1. Two-layer neural networks

We consider a two-layer neural network with N neurons in each layer. Thus the input and output patterns will be binary N-vectors; we take the symbols to be $\{1, -1\}$. The relationship between input pattern U and output pattern V is

$$V = \sigma(AU) \tag{1}$$

where $A = (a_{ij})$ is the connection matrix and σ is the sign function. Thus, given input patterns $U_1, U_2, \ldots, U_M$ and corresponding output patterns $V_1, V_2, \ldots, V_M$, where $U_m = (u_{mj})$ and $V_m = (v_{mj})$, the information obtained is

$$v_{mn} = \sigma\left(\sum_{j=1}^{N} a_{nj} u_{mj}\right), \quad 1 \le m \le M, \quad 1 \le n \le N. \tag{2}$$

The learning problem for this network is to find a connection matrix A satisfying (2), so each instance consists of a set of input patterns and the corresponding output patterns; thus an instance can be thought of as a mapping $U_m \rightarrow V_m : 1 \le m \le M$.

The solution may be unique, it may be nonexistent, or there may be a spectrum of possibilities for the a_{ij}. In the case of no solution we say the corresponding instance of the learning problem is *illegal*; otherwise it is *legal*.

[1]Research supported by a grant from the Neuroengineering Center at Southern Illinois University at Carbondale, Carbondale, IL.

2. Reduction of the problem (after [2])

Since $u_{mn} = \pm 1$, multiplying both sides of (2) by v_{mn} produces

$$1 = \sigma\left(\sum_{j=1}^{N} a_{nj}\, u_{mj} v_{mn}\right)$$

or equivalently

$$\sum_{j=1}^{N} a_{nj}\, w_{mnj} > 0\,, \quad 1 \le m \le M, \quad 1 \le n \le N \tag{3}$$

where the w_{mnj} are new variables defined by

$$w_{mnj} = u_{mj} v_{mn}. \tag{4}$$

If we denote by W_{mn} the (1,-1) vector of length N with j-th entry w_{mnj} (so that W_{mn} is column n of the matrix $U_m V_m^T$) and write A_n for row n of A, then (3) becomes

$$A_n W_{mn} > 0\,, \quad 1 \le m \le M, \quad 1 \le n \le N. \tag{5}$$

It is clear that the different inequalities, corresponding to the different values of n, do not interact. Therefore we have n instances of the following problem:

Find a vector A of length N such that

$$A W_m > 0\,, \quad 1 \le m \le M. \tag{6}$$

In the sequel we assume we are discussing (6). It will be convenient to write W for the matrix with m-th column W_m.

3. The acute cone case

In some cases one can assume that the vectors W_m always have positive inner product. In other words, the convex cone spanned by the vectors W_m is acute.

In this case, suppose A is any linear combination of the form

$$A = \sum c_i W_i^T$$

in which the c_i are non-negative and are not all zero (we call this a *positive linear combination* of the W). Then

$$A W_m = \sum c_i W_i^T W_m > 0$$

for all m. So we have the following rule, which is discussed in [2]:

Theorem 1. Geometric Learning Rule. *If the vectors W_{mn} for fixed n always form an acute convex cone C_n in N-space, then a solution A to the learning problem may be constructed by taking row n of A to be any positive linear combination of the W_{mn} (that is, any vector in C_n).*

4. The independent case

Suppose the vectors W_m are linearly independent. Then necessarily $M \leq N$, and we can find a matrix X such that if

$$S = [W : X]$$

then S is an $n \times n$ invertible matrix.

Suppose one wishes to attain a given set of values $h_1, h_2, \ldots, h_M$ for AW_1, $AW_2, \ldots, AW_M$. Write

$$H = (h_1, h_2, \ldots, h_M, 1, 1, \ldots, 1)$$

and use

$$A = HS^{-1}.$$

The values $h_1, h_2, \ldots, h_M$ can be chosen to be positive.

This case is of course not very interesting, but is included for completeness.

5. The general case

Suppose $A = X$ is a solution of (6). Let a be the value of the smallest member of XW_m. Then

$$(a^{-1}X)W_m \geq 1$$

(in the sense that every entry in the left-hand side is at least 1). So there is a solution to

$$AW_m \geq 1. \tag{7}$$

But obviously any solution of (7) also satisfies (6). So we have

Theorem 2. *To solve the learning problem for the network defined in (1), it is sufficient to find a solution of (7). If (7) is insoluble, then the learning problem is illegal.*

Unsolvable sets of inequalities were discussed by Kuhn [3]. A system

$$AM \geq b$$

is called *inconsistent* if and only if there exists a vector of nonnegative numbers Y such that

$$MY = 0$$

$$b.Y > 0.$$

In other words, (7) is inconsistent if and only if some vector Y of nonnegative numbers, not all zero, satisfies

$$MY = 0.$$

Theorem 3 [3; 1 p. 144]. *A system of linear inequalities is unsolvable if and only if it is inconsistent.*

Thus the above discussion gives a necessary and sufficient condition for the legality of the learning problem.

6. Computational considerations

Theoretical results aside, it is of interest to find a solution of (7), or prove that none exists, in specific cases.

The solution of a system such as (7) is no harder than solving linear programming problems; (7) is equivalent to

$$\begin{aligned} &\text{minimize} \quad a_0 \\ &\text{subject to} \quad -a_0 + AW_n \geq 1 \\ &a_0 \geq 0 . \end{aligned} \tag{8}$$

The learning problem is legal if and only if (8) has a non-empty solution set.

Accordingly, solution of the learning problem is reduced to solution of linear programming problems. This has been widely studied (see, for example, [1]), and many implementations are available.

References

1. V. Chvátal, in "Linear Programming", Freeman, New York, 1983.
2. C-L.J. Hu, *A novel geometrical supervised learning scheme*, SPIE 1294 Applications of Neural Networks (1990), 426–432.
3. H.W. Kuhn, *Solvability and consistency for linear equations and inequalities*, American Math. Monthly 63 (1956), 217–232.