

The Delta Rule

Developed by Widrow and Hoff, the delta rule, also called the Least Mean Square (LMS) method, is one of the most commonly used learning rules. For a given input vector, the output vector is compared to the correct answer. If the difference is zero, no learning takes place; otherwise, the weights are adjusted to reduce this difference. The change in weight from u_i to u_j is given by: $\Delta w_{ij} = r * a_i * e_j$, where r is the learning rate, a_i represents the activation of u_i and e_j is the difference between the expected output and the actual output of u_j . If the set of input patterns form a linearly independent set then arbitrary associations can be learned using the delta rule.

It has been shown that for networks with linear activation functions and with no hidden units (hidden units are found in networks with more than two layers), the error squared vs. the weight graph is a paraboloid in n -space. Since the proportionality constant is negative, the graph of such a function is concave upward and has a minimum value. The vertex of this paraboloid represents the point where the error is minimized. The weight vector corresponding to this point is then the ideal weight vector.

This learning rule not only moves the weight vector nearer to the ideal weight vector, it does so in the most efficient way. The delta rule implements a gradient descent by moving the weight vector from the point on the surface of the paraboloid down toward the lowest point, the vertex. Minsky and Papert raised good questions. Is there a simple learning rule that is guaranteed to work for all kinds of problems? Does the delta rule work in all cases?

As stated previously, it has been shown that in the case of linear activation functions where the network has no hidden units, the delta rule will always find the best set of weight vectors. On the other hand, that is not the case for hidden units. The error surface is not a paraboloid and so does not have a unique minimum point. There is no such powerful rule as the delta rule for networks with hidden units. There have been a number of theories in response to this problem. These include the generalized delta rule and the unsupervised competitive learning model.

This module on neural networks was written by Ingrid Russell of the University of Hartford. It is being printed with permission from Collegiate Microcomputer Journal.

If you have any comments or suggestions, please send an email to irussell@mail.hartford.edu
