

Strategies for Teaching Layered Networks Classification Tasks

Authors:

Ben Wittner, John Denker

Abstract:

There is a widespread misconception that the delta-rule is in some sense guaranteed to work on networks without hidden units.

As previous authors have mentioned, there is no such guarantee for classification tasks. We will begin by presenting explicit counterexamples illustrating two different interesting ways in which the delta rule can fail. We go on to provide conditions which do guarantee that gradient descent will successfully train networks without hidden units to perform two-category classification tasks. We discuss the

generalization of our ideas to networks with hidden units and to multi-category classification tasks. The

Classification Task Consider networks

of the form indicated in figure 1. We discuss various methods for training such a network, that is for adjusting its weight vector, w . If we call the input v , the output is $g(w \cdot v)$, where g is some function.

The classification task we wish to train the network to perform is the following. Given two finite sets of vectors, F_1 and F_2 , output a number greater than zero when a vector in F_1 is input, and output a number

less than zero when a vector in F_2 is input. Without significant loss of generality, we assume that g is

odd (i.e. $g(-s) = -g(s)$). In that case, the task can be reformulated as follows. Define F

$F := F_1 \cup \{-v \text{ such that } v \in F_2\}$ (1) and output a number greater than zero when a vector

in F is input. The former formulation is more natural in some sense, but the later formulation is somewhat more

convenient for analysis and is the one we use. We call vectors in F , training vectors. A Class of Gradient

Descent Algorithms We denote the solution set by $W := \{w \text{ such that } g(w \cdot v) > 0 \text{ for all } v$

$\in F\}$, (2) (Currently at NYNEX Science and Technology, 500 Westchester Ave., White Plains, NY 10604) We use both

$A := B$ and $A \doteq B$ to denote "A is by definition B". @ American Institute of Physics 1988