CONVERGENCE PROPERTIES OF A LEARNING ALGORITHM[1]

By Leo Breiman and Zivia S. Wurtele

University of California, Los Angeles

1. Introduction. In a recent paper Albert [1] introduced an algorithm for learning to classify individuals that are drawn from a population which is partitioned into two categories. The purpose of this note is to discuss an algorithm which is simpler, in the sense that at any given stage half as many items are retained in memory.

In the learning process described by the algorithm, observations are made on individuals one at a time and the current estimate of the required partitioning may be adjusted after each observation, on the basis of knowledge of the category to which the individual observed belongs. At any given time, the current estimate of the partitioning is all that is held in memory; past history is lost except insofar as it has been incorporated into the present estimate. The learning process of perceptrons, as well as that of other artificial intelligences, is of this general form.

2. Notation and assumptions. It is assumed that each individual is a member of one and only one of two categories. The results obtained are applicable to the more general case, however, for they may be applied to appropriate partitions of a set of three or more categories into two subsets.

Each individual in the population is characterized by an attribute vector X in m-dimensional Euclidean space; let S_1, S_2 be the sets of vectors attributed to members of the first and second categories, respectively. We shall suppose that this characterization is *sufficiently rich* with respect to the given classification problem, that is to say that the regions S_1 and S_2 are separable by a hyperplane (except for a set of probability measure zero). This terminology is appropriate to situations for which in the case of failure of the condition of sufficient richness, a re-examination of the world of individuals and the subsequent increasing of the number of components of the characterizing vectors can be expected to yield a new description for which this condition is satisfied. The question of whether, in a particular case, a sufficiently rich characterization can be achieved is obviously crucial but beyond the scope of this paper.

It is supposed that initially there are two samples: $X_1^{(1)}, \cdots, X_p^{(1)}$ from S_1 and $X_1^{(2)}, \cdots, X_q^{(2)}$ from S_2. Let X_n be the nth vector sampled after the initial $p + q$ vectors. We assume the following.

Assumption. There exist two bounded sets S_1, S_2 and a distribution Q on $S_1 \cup S_2$ such that

Received 31 October 1963; revised 8 May 1964.

[1] This work was supported partly by the Office of Naval Research under Task 047-041, and partly by the Western Management Science Institute under a grant from the Ford Foundation. Reproduction in whole or in part is permitted for any purpose of the United States Government.