

Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment

Authors: Michael C. Mozer, Paul Smolensky

Abstract: This paper proposes a means of using the knowledge in a network to determine the functionality or relevance of individual units, both for the purpose of understanding the network's behavior and improving its performance. The basic idea is to iteratively train the network to a certain performance criterion, compute a measure of relevance that identifies which input or hidden units are most critical to performance, and automatically trim the least relevant units. This skeletonization technique can be used to simplify networks by eliminating units that convey redundant information; to improve learning performance by first learning with spare hidden units and then trimming the unnecessary ones away, thereby constraining generalization; and to understand the behavior of networks in terms of minimal "rules."