What Size Net Gives Valid Generalization?

Authors: Eric Baum, David Haussler

Abstract: We address the question of when a network can be expected to generalize from m random training examples chosen from some ar(cid:173) bitrary probability distribution, assuming that future test examples are drawn from the same distribution. Among our results are the following bounds on appropriate sample vs. network size. Assume o < £ $ 1/8. We show that if m > O(~log~) random exam(cid:173) ples can be loaded on a feedforward network of linear threshold functions with N nodes and W weights, so that at least a fraction 1 - t of the examples are correctly classified, then one has confi(cid:173) dence approaching certainty that the network will correctly classify a fraction 1 - £ of future test examples drawn from the same dis(cid:173) tribution. Conversely, for fully-connected feedforward nets with one hidden layer, any learning algorithm using fewer than O('!') random training examples will, for some distributions of examples consistent with an appropriate weight choice, fail at least some fixed fraction of the time to find a weight choice that will correctly classify more than a 1 - £ fraction of the future test examples.