Approximation, complexity and risk properties of learning networks

Andrew Barron[a]

[a] *Department of Statistics and Data Science, Yale University, USA*

Abstract

For functions of bounded variation V with respect to a class of activation functions, we recall how an m term greedy approximation achieves error bounded by V/m^{1/2}, and we review its metric entropy and statistical risk implications for learning single hidden-layer neural networks. For certain high-dimensional inputs and deep networks, Gaussian and Rademacher complexity analysis are used to provide metric entropy and statistical risk bounds. The mean squared error of generalization is shown to have bounds of order V [(L+ log d)/n]^{1/2}, for networks with L layers, network variation V, input dimension d, and sample size n. These results show that favorable risk properties hold for extremely large numbers of input variables as long as the sample size n is large compared to the log of the number of variables.