Journal of Machine Learning Research X (2002) X-X Submitted 11/02; Published X/XX

CCIT Report # 420 March 2003

Greedy Algorithms for Classification - Consistency, Convergence Rates, and Adaptivity

Shie Mannor SHIE@MIT.EDU
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

Ron Meir RMEIR@EE.TECHNION.AC.IL
Department of Electrical Engineering
Technion, Haifa 32000, Israel

Tong Zhang TZHANG@WATSON.IBM.COM
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598

Editor: U.N. Known

Abstract

Many regression and classification algorithms proposed over the years can described as greedy procedures for the stagewise minimization of an appropriate cost function. Some examples include additive models, matching pursuit, and Boosting. In this work we focus on the classification problem, for which many recent algorithms have been proposed and applied successfully. For a specific regularized form of greedy stagewise optimization, we prove consistency of the approach under rather general conditions. Focusing on specific classes of problems we provide conditions under which our greedy procedure achieves the (nearly) minimax rate of convergence, implying that the procedure cannot be improved in a worst case setting. We also construct a fully adaptive procedure, which, without knowing the smoothness parameter of the decision boundary, converges at the same rate as if the smoothness parameter were known.

1. Introduction

The problem of binary classification plays an important role in the general theory of learning and estimation. While this problem is the simplest supervised learning problem one may envisage, there are still many open issues related to the best approach to solving it. In this paper we consider a family of algorithms based on a greedy stagewise minimization of an appropriate smooth loss function, and the construction of a composite classifier by combining simple base classifiers obtained by the stagewise procedure. Such procedures have been known for many years in the statistics literature as *additive models* (Hastie and Tibshirani, 1990, Hastie et al., 2001) and have also been used in the signal processing community under the title of *matching pursuit* (Mallat and Zhang, 1993). More recently,