Mach Learn (2012) 86:391–423
DOI 10.1007/s10994-011-5268-1

Robustness and generalization

Huan Xu · Shie Mannor

Received: 16 February 2011 / Accepted: 10 October 2011 / Published online: 15 November 2011

Abstract We derive generalization bounds for learning algorithms based on their robustness: the property that if a testing sample is "similar" to a training sample, then the testing error is close to the training error. This provides a novel approach, different from complexity or stability arguments, to study generalization of learning algorithms. One advantage of the robustness approach, compared to previous methods, is the geometric intuition it conveys. Consequently, robustness-based analysis is easy to extend to learning in non-standard setups such as Markovian samples or quantile loss. We further show that a weak notion of robustness is both sufficient and necessary for generalizability, which implies that robustness is a fundamental property that is required for learning algorithms to work.

Keywords Generalization · Robust · Non-IID sample · Quantile loss

1 Introduction

The key issue in the task of learning from a set of observed samples is the estimation of the *risk* (i.e., generalization error) of learning algorithms. Typically, its empirical measurement (i.e., training error) provides an optimistically biased estimation, especially when the number of training samples is small. Several approaches have been proposed to bound the deviation of the risk from its empirical measurement, among which methods based on uniform convergence and stability are most widely used.

Uniform convergence of empirical quantities to their mean (Vapnik and Chervonenkis 1974, 1991) provides ways to bound the gap between the expected risk and the empirical risk by the complexity of the hypothesis set. Examples of complexity measures are

Editor: Phil Long.

H. Xu (✉)
Department of Mechanical Engineering, National University of Singapore, Singapore, Singapore
e-mail: mpexuh@nus.edu.sg

S. Mannor
Department of Electrical Engineering, Technion, Israel Institute of Technology, Haifa, Israel
e-mail: shie@ee.technion.ac.il