ROC convex hull and nonparametric maximum likelihood estimation

Johan Lim · Joong-Ho Won

Abstract The ROC convex hull (ROCCH) is the least convex majorant of the empirical ROC curve, and represents the optimal ROC curve of a set of classifiers. This paper provides a probabilistic view to the ROCCH. We show that the ROCCH can be characterized as a nonparametric maximum likelihood estimator (NPMLE) of a convex ROC curve. We provide two NPMLE formulations, one unconditional and the other conditional, both of which yield the ROOCH as the solution. The solution technique relates the NPMLEs to convex optimization and classifier calibration. The connection between the NPMLEs and the ROCCH also suggests efficient algorithms to compute NPMLEs of a convex ROC curve, and a conditional bootstrap procedure for assessing uncertainties in the ROCCH.

1 Introduction

A receiver operating characteristic (ROC) curve is a graphical representation of two performance measures of binary classifiers, the false positive rate (FPR) and the true positive rate (TPR). The FPR is the probability of erroneously reporting negative instances as being positive, whereas the TPR is that of correctly reporting positive instances. The ROC space is a set of (FPR, TPR) pairs. Traditionally the ROC space is visualized by plotting the FPR on the x axis and the TPR on the y axis. A classifier that reports a class label corresponds to a point in the ROC space. To be specific, suppose a diagnostic test uses a continuous

Editor: Carla Brodley.

J. Lim
Department of Statistics, Seoul National University, Seoul, Korea
e-mail: johanlim@snu.ac.kr

J.-H. Won (✉)
VA Cooperative Studies Program, Mountain View, CA 94043, USA
e-mail: JoongHoJohann.Won@va.gov