A boosting method for maximization of the area under the ROC curve

Osamu Komori

Abstract We discuss receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) for binary classification problems in clinical fields. We propose a statistical method for combining multiple feature variables, based on a boosting algorithm for maximization of the AUC. In this iterative procedure, various simple classifiers that consist of the feature variables are combined flexibly into a single strong classifier. We consider a regularization to prevent overfitting to data in the algorithm using a penalty term for nonsmoothness. This regularization method not only improves the classification performance but also helps us to get a clearer understanding about how each feature variable is related to the binary outcome variable. We demonstrate the usefulness of score plots constructed componentwise by the boosting method. We describe two simulation studies and a real data analysis in order to illustrate the utility of our method.