A General AUC-type Measure and Biomarkers Selection

張源俊

中央研究院統計科學研究所

Abstract

Receiver operating characteristic (ROC) curve-based methods are well-accepted measures for describing and comparing the accuracy of diagnostic tests and biomarkers, which are essential in, for example, medical diagnoses, and development of drugs and treatments. However, conventional ROC curve-based methods depend on the existence of a binary-scale gold standard and cannot be directly applied to a case with only a continuous gold standard and no binary references. One possible approach is to dichotomize such a continuous gold standard into a binary surrogate. However, to do so, we have to pay the price for the dubious dichotomization. In order to overcome this shortcoming, we propose a novel area under ROC curve-type measure, such that one can measure the association between biomarkers or diagnostic tests with the continuous gold standard directly, which extends the traditional area under the ROC curve and includes a measure reported in the literature as its special case. Moreover, it is well known that the diagnostic power of an individual marker can be improved through the combination of multiple potential markers. Hence, in this paper, we also propose several algorithms to nd the best linear combination of markers, with respect to this newly proposed measure. The asymptotic properties of both estimates of this new measure and of the best combination vector are studied. We then illustrate the performance of the proposed method using both synthesized and real datasets.