Selecting Markers in Binary Classification Problems through their Linear Scores that Maximizes the (partial) Area under Receiver Operating Characteristic Curve

Yuan-Chin Chang, Academia Sinica

Abstract

Rather than viewing receiver operating characteristic (ROC) curves directly for comparing the performances of diagnostic methods, the whole and the partial areas under ROC curve, say AUC and pAUC, are two most popularly used summaries of the curve among others. The threshold independence of AUC is one of its advantages, On the other hand, when high specificity is a prerequisite as in some medical diagnostic, the pAUC is usually preferable to others. In Su and Liu (1993), they studied the best linear combination of markers that maximizes the AUC. Their idea can be extended for the pAUC as well. But these methods rely on normality assumption and may suffer from high dimensional data, which is a very common situation in the modern high throughput experiments. In this talk, we discuss the marker selection through finding linear combination of markers that maximizing the AUC or partial AUC. Some algorithms based on nonparametric estimation techniques are proposed for dealing with data set with lengthy markers.