Summary of doctoral thesis

Nonparametric methods of the ROC Curve estimation

Michał Pulit

The receiver operating characteristic (ROC) curve is a graphical representation of the relationship between false positive and true positive rates. It is a widely used statistical tool for describing the accuracy of a diagnostic test. Its application involve e.g. medicine, economy, engineering, psychology and machine learning. In the dissertation we take the problem of nonparametric estimation of the ROC curve.

We consider three different types of methods of nonparametric estimation of the ROC curve: plug-in methods, methods based on the fact that the ROC curve is a cumulative distribution function of some random variable and methods involving regression models. In two of the three cases, we propose new estimators of the ROC curve.

In the chapter related to pulg-in methods, we propose an estimator of the ROC curve based on the smoothed empirical distribution functions. We prove some properties of the estimator including its strong consistency.

In the chapter devoted to the methods based on the idea of estimating the ROC curve as a distribution function, we propose two other nonparametric estimators of the ROC curve. We show some properties of both estimators, e.g. their invariance under nondecreasing data transformations. Moreover, in the case of one of them, involving kernel smoothing method, we prove that the estimator has less asymptotic mean squared error than some other popular nonparametric estimators of the ROC curve. We also present an easy method of bandwidth selection for this estimator.

In order to compare the performance of the proposed estimators of the ROC curve and illustrate their advantage over some other estimators when sample sizes are small, we conducted several simulation studies. Their results are presented in the last chapter.