Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:5**
http://www.math.su.se

Predictive Performance of AdaBoost and Random Forest in Binary Classification Tasks

Markus Söderqvist*

May 2024

Abstract

Binary classification is the task of classifying an observation into one of two classes. In this thesis we compare the predictive performance of two machine learning algorithms for binary classification, AdaBoost and random forest. We do so on nonlinear data sets, where nonlinearity is achieved by enclosing one class in a geometrical shape in the predictor space. The comparisons are conducted on data sets with (i) noise, (ii) skewed class distribution and (iii) redundant predictors. In addition, we investigate how predictor dimension affects performance. Overall, we find that the two methods have similar performance, although some differences emerge. Random forest has a higher accuracy on noisy data sets, while AdaBoost has a higher accuracy on data sets with skewed class distribution and redundant predictors. Moreover, AdaBoost tends to outperform random forest on data sets with higher predictor dimension. The cost of this advantage is a considerably longer runtime. These findings are in line with previously reported findings. One unexpected finding is that the performances of both methods improve when the class distribution is skewed. A further analysis shows that one class is easier to classify at the expense of the other class for skewed data sets. Therefore, one should be careful about drawing conclusions from these results. Finally, an in-depth analysis in higher predictor dimensions shows that random forest has superior accuracy on one class while AdaBoost has superior accuracy on the other class. One possible explanation could be how the algorithms are constructed, and this can have important implications for choice of method in other classification problems.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: soderqvist.markus@gmail.com. Supervisor: Ola Hössjer and Johannes Heiny.