

Comparison of Different Classification Methods on Glass Identification for Forensic Research

Suchismita Goswami^{1,*} and Edward J. Wegman¹

1. Computational and Data Sciences, George Mason University, Fairfax VA 22030

*Contact author : sgoswam3@masonlive.gmu.edu

Keywords: Forensic Research, PCA, QDA, LDA, Classification Tree

Abstract

Classification methods play an important role in investigating crime in forensic research. Here we assessed relative performance of several classification methods, such as Logistic Regression (LR), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Mixture Discriminant Analysis (MDA) and Classification Tree (CT) on glass identification data. The area under the receiver operating characteristic curve (AUC), error rate and its 95% confidence interval were used to measure predictive power of these algorithms. Parallel coordinate (PC) plots were also introduced to characterize the data. Dimensionality reduction of data was conducted using principal component analysis (PCA) and Fisher's linear discriminant analysis (FDA) and two major components were identified. Among all the classification methods mentioned above, the LDA and the QDA were observed to be statistically significant. The Box's M test ($P < 0.0001$), which is used to test the homogeneity of covariance matrices showed that the homogeneity of covariance could not be assumed for LDA. This suggests that for glass types, window and non-window, the QDA is superior to all methods. The CT, however, outperforms all methods when all six categories of glass were considered. Also, it was observed that the tour-prune strategy in parallel coordinate plots resulted in a decision tree similar to the decision tree obtained by the classification tree method.