

REAL-TIME DISCRIMINANT ANALYSIS IN THE PRESENCE OF LABEL AND MEASUREMENT NOISE

Mia Hubert¹, Iwein Vranckx¹, Jakob Raymaekers², Bart De Ketelaere³ and Peter Rousseeuw¹

¹ Section of Statistics and Data Science, Department of Mathematics, KU Leuven (e-mail: mia.hubert@kuleuven.be, peter@rousseeuw.net)

² Department of Quantitative Economics, Maastricht University, (e-mail: j.raymaekers@maastrichtuniversity.nl)

³ Division of Mechatronics, Biostatistics and Sensors, (e-mail: bart.deketelaere@kuleuven.be)

ABSTRACT: Quadratic discriminant analysis (QDA) is a widely used classification technique. Based on a training dataset, each class in the data is characterized by an estimate of its center and shape, which can then be used to assign unseen observations to one of the classes. The traditional QDA rule relies on the empirical mean and covariance matrix. Unfortunately, these estimators are sensitive to label and measurement noise which often impairs the model's predictive ability. Robust estimators of location and scatter are resistant to this type of contamination. However, they have a prohibitive computational cost for large scale industrial experiments. We present a novel QDA method based on a real-time robust algorithm. We additionally integrate an anomaly detection step to classify the most atypical observations into a separate class of outliers. Finally, we introduce the classmap, a graphical display to identify label and measurement noise in the training data.

KEYWORDS: minimum covariance determinant, mislabeling, outliers, robust classification