

Linear Discriminant Analysis: Applications and Challenges in Data Analysis

Student: Turcu Ciprian-Stelian

Master: High Performance Computing and Big Data Analytics

E-mail: ciprian.turcu@stud.ubbcluj.ro

This study discusses the use of Linear Discriminant Analysis (LDA) in dimensionality reduction and enhancement of class separability on Wine Quality, Mall Customers, Heart Disease, Diabetes, and Online Retail datasets. It shows how LDA is effective in reducing data complexity while preserving important patterns, thus being useful for classification and interpretability.

In healthcare, LDA proved useful in identifying significant differentiators of cholesterol and glucose levels that distinguished patients with heart disease or diabetes from healthy individuals. LDA enhances the separability in health data, therefore supporting diagnostics with precision and targeted risk prevention. In wine quality analysis, LDA underlined such chemical attributes as alcohol and acidity as significant in making a difference in quality and provided proper insight into the refinement process to maintain the standard of a product.

Its application to customer behaviour analysis had more spending and demographic patterns that allowed it to segment shoppers of malls and online retail platforms. The analytics hence allowed the creation of customized marketing policies and supply chains, particularly relating to the identification of high-value segments of customers.

The study further underlines several theoretical strengths of LDA, such as reducing high-dimensional data without necessarily sacrificing class separability. However, it flags challenges in that LDA is sensitive to violations in the assumptions of multivariate normality and equal covariance and thus may not perform as well on a dataset.

Overall, the findings confirm that LDA is indeed an effective but interpretive dimensional data reduction and classification technique. Even though Quadratic Discriminant Analysis (QDA) and Regularized Discriminant Analysis (RDA) fit better when the data fails on most of the assumptions required for LDA, its ease of computation and efficiency make it invaluable in many practical fields including healthcare, product quality analytics, and consumers' behaviour.