

Mathematical Statistics Stockholm University Bachelor Thesis **2019:10** http://www.math.su.se

A Comparative Simulation Study of Logistic Regression and Linear Discriminant Analysis for Classification

Sofie Jörgensen*

June 2019

Abstract

The process of learning from data is central in statistical learning and the problem of selecting an appropriate method for a particular situation can be rather challenging. Two widely used linear classification methods are logistic regression and linear discriminant analysis. This thesis aims to compare these linear classifiers, partly from a theoretical perspective and partly through practical simulations, in order to study their similarities and differences in several aspects. The theoretical part outlines the concept of statistical learning and provides a detailed presentation of the methods of interest. The simulation part contains four experiments with different setups in order to evaluate the predictive power for each method. It turned out that logistic regression and linear discriminant analysis performed similarly, despite the fact that a variety of simulated data sets were used. Some notable differences were observed, which can be explained by the two methods' different ways of estimating parameters. Overall, the simulation study agreed with the theory provided in this thesis, that the two methods give similar results but that their prediction accuracy might deviate slightly from each other in some situations. This emphasizes the importance of examining the underlying structure of data before determining which method to use.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: jorgensen.sof@gmail.com. Supervisor: Ola Hössjer.