

Mathematical Statistics Stockholm University Bachelor Thesis **2018:20** http://www.math.su.se

Logistic Regression versus Support Vector Machines

Alexander Nöu*

June 2018

Abstract

In binary classification, the objective is to identify if an observation belongs to one of two classes, on the basis of data consisting of observations with known classes. Logistic regression and support vector machines are two of several classification methods, where the former is a traditional statistics method, and the latter is part of the closely related field of machine learning. The purpose of this thesis is to analyze and compare these methods, both theoretically and practically.

In the theoretical part, each method is described in detail, from concept to model fitting. Despite their different approaches to the classification problem, it turns out that the parameters for both methods can be obtained by minimizing an objective function, consisting of a loss and penalty function. The logistic regression loss and the support vector machine loss behave similarly, and as a result, they often have similar prediction accuracies.

The practical part consists of four different experiments, where in each experiment, a number of binary data sets are simulated. The aim is to analyze how each classifier performs on different types of data, by varying a number of parameters that characterize a data set.

The results coincided with the theory, showing that the predictive power of logistic regression and the support vector machine was close to equal (a few minor differences were observed). It was also shown that each method can be modified in order to obtain some of the advantages of the other method, making them even more alike.

^{*}Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: alexander.lacson.nou@gmail.com. Supervisor: Ola Hössjer and Disa Hansson.