Mathematical Statistics
Stockholm University
Bachelor Thesis **2024:2**
http://www.math.su.se

A comparative study of binary classification performance with Logistic Regression, Support Vector Machines and Artificial Neural Networks

Liliya Trila*

January 2024

Abstract

Binary classification is a common task in machine learning, where the goal is to categorize data into one of two classes. In this thesis we compare three methods for binary classification: logistic regression, support vector machines and artificial neural networks. The aim of this project is to understand the similarities and differences between these methods.

In the first part of the project, we present the theory related to each method. We also do a theoretical comparison of the methods.

In the second part of this project we compare these methods on both simulated and real-world data. To evaluate the performance of each method we use the performance metrics accuracy and AUC. The results show that logistic regression and support vector classifiers have very similar performance when the data is linearly separable. The performance of artificial neural networks tend to be slightly lower for small and high-dimensional datasets. When the data is not linearly separable artificial neural network models tend to perform slightly better on lower-dimensional data while support vector machines with an RBF kernel perform slightly better on higher-dimensional data. But overall, their performance is quite comparable.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: liliatrila@gmail.com. Supervisor: Taras Bodnar and Johannes Heiny.