With R Programming, Comparison of Performance of Different Machine Learning Algorithms

Prof. Dr. Ayşe OĞUZLAR

Uludağ Üniversitesi

Yusuf Murat KIZILKAYA

Ardahan Üniversitesi

Abstract

Machine Learning (ML) includes automatic calculation procedures based on logical or binary operations that learn a set of tasks. There are statistical approaches in the background of ML's decision process. ML uses statistical theory to construct mathematical models, because the main task is to inference by a set of data. ML programs computers in order to optimize a process based on past experience and / or example data. By ML, desired classifications can be done by computer in a short time and effectively. A model is created and this model can be foreshadowed in future predictions, can be found in explanations, or can be inspected on the basis of available data. ML functions in three different ways. The first is supervised learning, the second is unsupervised learning, and the third is semi-supervised learning. In supervised learning used in this study; a set of data and a training set about the concept to be learned in the system is entered. In the training set, the desired output values are also given for each sample (labeling done). Moving from this information, a relationship is established between input and output. Output values are tried to be estimated or learned by using the values of the input data. The results are classified based on known data and predictions are made on data sets whose results are unknown. In this study, in R programming, machine learning performances are compared. For this purpose, various machine learning algorithms have been applied to real data obtained from the UCI machine learning repository which is a collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms, (UCI was created as an ftp archive in 1987 by David Aha and fellow graduate students at UC Irvine. Since that time, it has been widely used by students, educators, and researchers all over the world as a primary source of machine learning data sets), and classification algorithms have been compared using various criteria. The calculated criteria are; precision, accuracy, sensitivity, and classifi- cation techniques based on the F-scale were compared. As a result of comparisons made, Logistic Regression algorithm is seen that to be more successful than other algorithms.

Keywords: R Programing, Machine Learning, Survised Learning, Classification Algorithms.