Chapter 4
Introduction and Implementation of Machine Learning Algorithms in R

S. R. Mani Sekhar
M. S. Ramaiah Institute of Technology, India

G. M. Siddesh
M. S. Ramaiah Institute of Technology, India

ABSTRACT

Machine learning is one of the important areas in the field of computer science. It helps to provide an optimized solution for the real-world problems by using past knowledge or previous experience data. There are different types of machine learning algorithms present in computer science. This chapter provides the overview of some selected machine learning algorithms such as linear regression, linear discriminant analysis, support vector machine, naive Bayes classifier, neural networks, and decision trees. Each of these methods is illustrated in detail with an example and R code, which in turn assists the reader to generate their own solutions for the given problems.

INTRODUCTION

In this age, the advancements in the arena of Machine learning are huge. Machine Learning has become one emerging field of computer science. As a part of many different algorithms that can learn form and make decisions on data are developed. Different algorithm works well in different situation. There are different machine learning application have been developed such as algorithms to filter e-mails, fraud detection, voice recognition, and weather prediction etc.

This chapter provides an overview of some machine learning algorithms such as Linear Regression, Linear Discriminant Analysis, Support Vector Machine, Naive Bayes classifier, Neural Networks, & Decision trees. Thereafter each of these algorithms is explained with respect to its use, application and supported examples. Finally for each of the above stated algorithm a detailed line by line R Code explanation is provided.

DOI: 10.4018/978-1-7998-7705-9.ch004

The rest of the chapter is organized as follows: Section 2, discuss the details about the linear Regression with supported example and R code. Section 3, tells the details of Linear Discriminant Analysis, LDA classifier, dimensionality reduction with supported R Codes. Subsequently Section 4, discuss the support vector machine and its code. Section 5, illustrates Naive Bayes Classifier with supported application and R Code. Furthermore Section 6, Discuss Neural Network implementation, working, application and its advantages. Later Section 7 discusses the concept of Decision trees, it implementation. Finally section 8 discusses the conclusion part.

LINEAR REGRESSION

Linear regression (Kenney & Keeping, 1962) is a Machine learning method that is used to model the relationship between the two variables one of which is called explanatory variable or independent variable and the other one is called target variable or dependent variable.

If there is only one explanatory variable then the problem is called *simple linear regression* whereas if there are more explanatory variable then the problem is called *multiple linear regression.*

Mathematically, the relationship modeled using linear regression approach yields a straight line when plotted as a graph. Conventionally, the explanatory variable is represented by X and target variable is represented by Y.

The general mathematical equation for linear regression (Kenney & Keeping, 1962) is given by,

$$Y = AX + B \text{ ; where A \& B are constants.} \quad (1)$$

Linear regression is used when there is a some set of known pairs of X and Y values and we have to predict new Y value for every new X value using the predefined model as shown in equation 1. The widely used approach to fit the linear regression model is "*Least Square Method*" (Zhou & Han, 1951). Linear regression approach is used whenever there is a task of modeling or analysis of several variables.

The real life example of the linear regression is predicting the weight of the person whose height is known or else predicting the distance that the car will travel whose speed is known or predicting the weather where temperature is known.

Consider the example for predict the weight of the person whose height is known. For this, first collect the sample data where the heights and weights of the person are mentioned. Using the collected data, generate a relationship model by applying linear regression. Simultaneously find out the value of the constants from the model created. Thereafter plot the graph and obtain the line which is always straight. Now, the model can be used to predict the weight for the given person height. Table 1, show the sample dataset named "*trees*"(Ryan, Joiner, & Ryan,1976) it is a build-in dataset of RStudio consist of 4 columns, which represent the Girth, Height and volume of trees.

Working Algorithm Code

The linear regression can be used to model the relationships between X and Y variable easily using R. Though there is an already available function in R which performs linear regression (i.e. lm() function), here the aim is to develop a linear regression model from scratch without using the built-in function us-