**Experiment 9**

|  |  |
| --- | --- |
| **Student Name: Nikhil Kumar** | **UID: 20BCS1817** |
| **Branch: BE-CSE** | **Section/Group:716/B** |
| **Semester: 6th** | **Date of Performance: 09/05/2023** |
| **Subject Name: Data Mining Lab** | **Subject Code: 20CSP-376** |

1. **Aim:** Study of Regression Analysis using R Programming.

1. **Objective:** Linear Regression: It is a commonly used type of predictive analysis.

It is a statistical approach for modelling the relationship between a dependent variable and a given set of independent variables.

There are two types of linear regression.

* + Simple Linear Regression
  + Multiple Linear Regression

**Simple Linear regression using R.**

It is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables. One variable denoted x is regarded as an independent variable and the other one denoted y is regarded as a dependent variable. It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value as accurately as possible as a function of the feature or independent variable(x).

1. **Script and Output:**

The algorithm is as follows:

* + Generates 40 random IQ values with a mean of 30 and a standard deviation of

2 and assign to “**IQ**” vector.

* + Sorts the **IQ** vector in ascending order.
  + Creates a vector **result** that contains pass (1) and fail (0) values for the 40 students.
  + Combing the **IQ** and **result** vectors using **cbind()** and then converting them to a data frame **df** using **as.data.frame().**
  + Opens a PNG file named “**LogisticRegressionGFG.png**” for writing.
  + Creates a scatter plot of the **IQ** and **result** variables.
  + Fits a logistic regression model to the data.

**R Script:**

# Generate random IQ values with mean = 30 and sd =2

IQ <- rnorm(40, 30, 2)

# Sorting IQ level in ascending order

IQ <- sort(IQ)

# Generate vector with pass and fail values of 40 students result <- c(0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0,

0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1)

# Data Frame

df <- as.data.frame(cbind(IQ, result))

# Print data frame print(df)

# output to be present as PNG file png(file="LogisticRegressionGFG.png")

# Plotting IQ on x-axis and result on y-axis plot(IQ, result, xlab = "IQ Level",

ylab = "Probability of Passing")

# Create a logistic model g = glm(result~IQ, family=binomial, df)

# Create a curve based on prediction using the regression model

curve(predict(g, data.frame(IQ=x), type="resp"), add=TRUE)

# This Draws a set of points

# Based on fit to the regression model points(IQ,

fitted(g), pch=30)

# Summary of the regression model summary(g)

# saving the file dev.off()

**Output:**





