

Practical 10: Perform the logistic regression on the given data warehouse data.

<https://rdr.io/snippets/>

Solution:

Multiple linear regression:

It is a machine learning technique used to predict the value of a dependent variable (here it is Admitted) based on two or more independent variables(here hsc,ssc,cet and viva). In other words, it helps to understand how different variables affect the outcome of something. **For example**, it can be used to predict a person's salary based on their age, education, and experience. The model estimates the relationship between the variables and uses this to make predictions. **It is widely used in various fields such as finance, economics, engineering, and social sciences.**

Following is a R code to determine whether a student will be admitted based on their marks in HSC, CET, and Viva. The total number of students is 70. The marks in each subject are out of 100. The dependent variable, 'Admitted', should be binary with a value of 1 indicating 'yes' and 0 indicating 'no'.

```
#Get Marks of 3 subjects, viva and status of admission
```

```
hsc <-
```

```
c(55,26,69,98,98,88,76,99,62,56,87,99,84,25,68,65,46,58,82,93,34,65,57,79,65,34,51,32,15,23,22,11,79,65,34,51,32,15,23,22,11,10,57,59,64,65,22,23,23,45,4,8,6,35,35,36,39,35,36,68,69,64,35,34,34,56,26,35,34,35)
```

```
ssc <-
```

```
c(32,35,31,24,68,65,46,58,82,93,34,99,84,25,31,65,24,53,39,32,35,31,24,68,65,46,58,82,93,34,65,57,79,65,34,51,32,15,23,22,11,10,57,59,64,65,22,23,23,45,4,5,6,56,65,65,35,34,26,69,68,98,69,68,69,78,45,65,68,69)
```

```
cet <-
```

```
c(55,26,69,98,98,88,76,99,62,56,87,99,84,25,31,65,24,53,39,32,35,31,24,68,65,46,58,82,93,34,65,57,79,65,34,51,32,15,23,22,11,10,57,59,64,65,22,23,23,45,3,8,2,59,68,68,54,58,69,65,69,69,65,58,57,59,69,68,58,65)
```

```
viva <-
```

```
c(68,65,46,58,82,93,34,65,57,79,65,34,51,32,15,23,24,53,39,32,35,31,24,68,65,46,58,82,93,34,65,57,79,65,34,51,32,15,23,22,34,51,32,15,23,22,11,10,57,59,5,8,6,78,98,54,69,68,98,58,69,58,54,45,65,56,44,98,45,54)
```

```
admitted <-
```

```
c(0,0,0,0,1,1,0,1,1,1,0,0,1,0,0,0,0,1,1,0,0,0,0,1,1,0,1,0,0,0,0,0,1,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,0,1,0,0,0,0,1,1,1,0,0,0,1,0,0,0,0)
```

Create a data frame with the independent and dependent variables

```
data <- data.frame(hsc, ssc, cet, viva, admitted)
```

Fit the logistic regression model

```
logit_model <- glm(admitted ~ hsc + ssc + cet + viva, data = data, family = "binomial")
```

Print the model summary

```
summary(logit_model)
```

Predict the admission status for a new student

```
new_student <- data.frame(hsc = 67, ssc = 98, cet = 86, viva = 74)
```

```
predict(logit_model, new_student, type = "response")
```

Output

```
1  
0.9502889
```

This output suggests a predicted probability of admission for the new student, calculated as approximately 95.03%.