CS 4407-01 Data Mining and Machine Learning

Instructor: Professor Shabia Shabir

Name: Anonymos

Programming Assignment Unit 5

**Part 1: Output of Decision Tree**

The decision tree for the given Ionosphere dataset is constructed and visualized using R. The steps are as follows:

**a. Setting the Working Directory and Loading Packages**

Set the working directory, load the required packages (`rpart` and `mlbench`), and then load the Ionosphere dataset.

```R

# Setting the working directory (modify as needed)

setwd("/Users/hayashir/work/Github/UoPeople/CS4407-branch/CS4407/unit5/PA")

# Loading packages

library(rpart)

library(mlbench)

# Loading the dataset

ionosphere\_data <- read.table("Ionosphere.txt",

header = FALSE,

sep = ",",

stringsAsFactors = TRUE)

# Adding column names

colnames(ionosphere\_data) <- c(paste0("V", 1:34), "Class")

# Verifying the dataset

head(ionosphere\_data)

summary(ionosphere\_data)

```

This step ensures that the Ionosphere.txt dataset is correctly loaded as a dataframe with 34 continuous columns and 1 class label column (`Class: "g" or "b"`).

**b. Creating the Decision Tree**

Use the `rpart()` function to create a decision tree for classification. This model predicts the `Class` (whether the signal is good or bad).

```R

# Creating the decision tree

ionosphere\_tree <- rpart(Class ~ ., data = ionosphere\_data)

# Displaying the summary of the decision tree

print(summary(ionosphere\_tree))

```

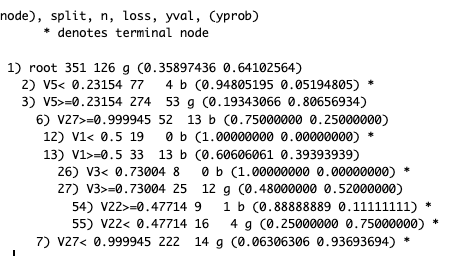
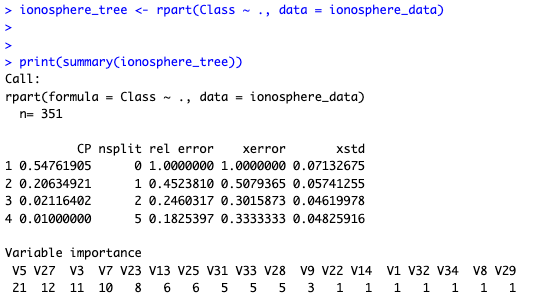


Figure 1: Overview of the Decision Tree

**c. Visualizing the Decision Tree**

Use the `plot()` and `text()` functions to visualize the constructed decision tree. The commands are as follows:

```R

# Adjusting margins

par(mar = c(4, 4, 4, 4))

# Plotting the decision tree

plot(ionosphere\_tree, uniform = TRUE, main = "Ionosphere Decision Tree")

# Adding labels to nodes

text(ionosphere\_tree, use.n = TRUE, all = TRUE, cex = 0.8)

```

The results are as follows:

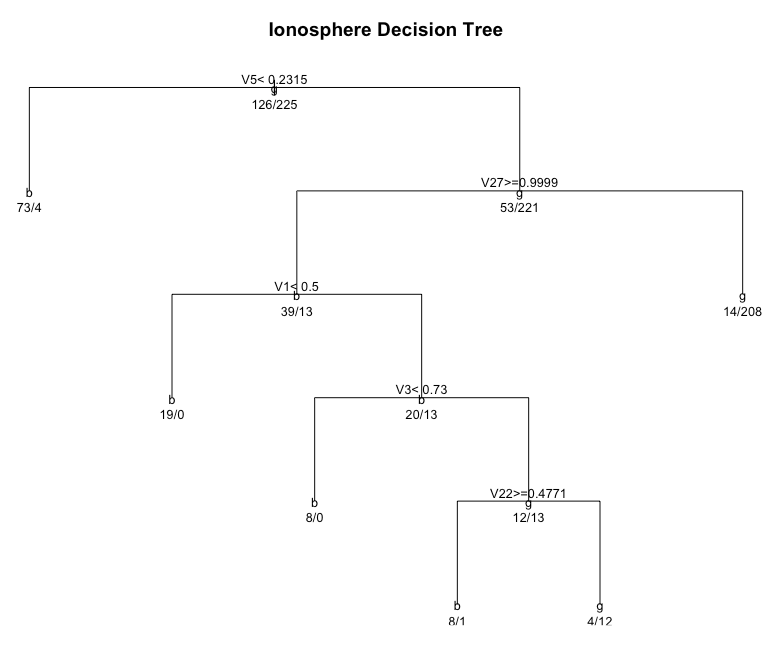


Figure 2 Ionosphere Decision Tree

**Part 2: Evaluating Accuracy**

This section involves splitting the Ionosphere dataset into training and testing datasets, building a decision tree, and evaluating its prediction accuracy on the test data.

**a. Splitting the Data**

Split the dataset into training (70%) and testing (30%) datasets using the `sample()` function.

```R

set.seed(123)

# Splitting the data into 7:3

train\_indices <- sample(1:nrow(ionosphere\_data), 0.7 \* nrow(ionosphere\_data))

train\_data <- ionosphere\_data[train\_indices, ]

test\_data <- ionosphere\_data[-train\_indices, ]

```

**b. Creating the Decision Tree with Training Data**

Build a new decision tree using the training data.

```R

# Creating the decision tree with training data

ionosphere\_tree\_train <- rpart(Class ~ ., data = train\_data)

# Displaying the summary of the model

print(summary(ionosphere\_tree\_train))

```

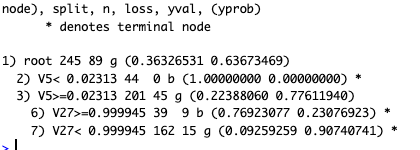
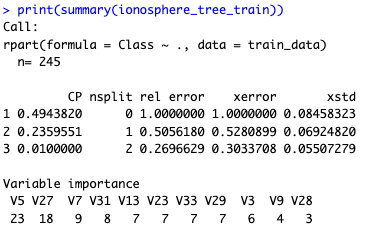


Figure 3: Decision Tree Overview for Test Data

**c. Predicting with the Test Data**

Use the `predict()` function to make predictions based on the test data.

```R

# Making predictions with the test data

test\_predictions <- predict(ionosphere\_tree\_train, newdata = test\_data, type = "class")

```

**d. Calculating Accuracy**

Compare the predicted results with the actual class labels in the test data and calculate the model's accuracy.

```R

# Creating a confusion matrix

confusion\_matrix <- table(test\_predictions, test\_data$Class)

# Calculating accuracy

accuracy <- sum(diag(confusion\_matrix)) / sum(confusion\_matrix)

# Displaying the results

print(confusion\_matrix)

print(paste("Accuracy: ", round(accuracy \* 100, 2), "%"))

```

The measured accuracy is as follows:

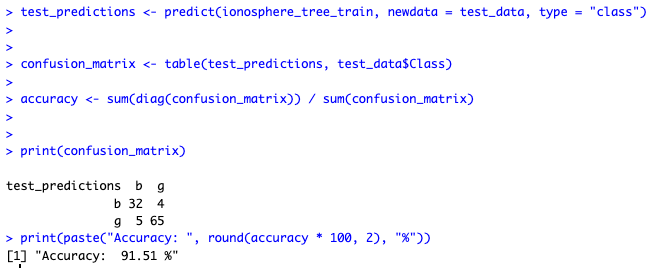


Figure 4: Accuracy Calculation of the Model

Word Count: 476

References

1. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York, NY: Springer.