Complete R Programming for Machine Learning Cheatsheet

Introduction

This cheatsheet provides a comprehensive overview of R programming concepts for machine learning. It is designed to assist beginners in understanding and implementing machine learning algorithms in R.

Installation

To get started with R and machine learning packages, follow these steps:

```r
# Install R
install.packages("r-base")

# Install required packages
install.packages(c("tidyverse", "caret", "randomForest"))
```

Data Manipulation

R provides powerful tools for data manipulation and preprocessing.

Loading Data

To load a dataset in R, use the following code:

```r
# Load CSV data
data <- read.csv("data.csv")

# Load Excel data
library(readxl)
data <- read_excel("data.xlsx")
```

Data Exploration

R offers various functions for exploring and summarizing data.

```r
# Summary statistics
summary(data)

# Correlation matrix
cor(data)

# Frequency table
table(data$column)
```

Machine Learning Algorithms

Linear Regression

To perform linear regression in R:

```
1  # Load the caret package
2  library(caret)
3
4  # Split data into training and testing sets
5  set.seed(123)
6  trainIndex <- createDataPartition(data$target, p = 0.8, list = FALSE)
7  trainData <- data[trainIndex, ]
8  testData <- data[-trainIndex, ]
9
10 # Train the linear regression model
11 lm_model <- train(target ~ ., data = trainData, method = "lm")
12
13 # Make predictions on test data
14 predictions <- predict(lm_model, newdata = testData)
```

Decision Trees

To build a decision tree in R:

```
1  # Load the rpart package
2  library(rpart)
3
4  # Build the decision tree
5  tree_model <- rpart(target ~ ., data = trainData, method = "class")
6
7  # Plot the decision tree
8  plot(tree_model)
9  \end
10 \subsection*{Random Forest}
11 To build a random forest model in R:
12
13 \begin{lstlisting}[language=R]
14 # Load the randomForest package
15 library(randomForest)
16
17 # Build the random forest model
18 rf_model <- randomForest(target ~ ., data = trainData, ntree = 100)
19
20 # Make predictions on test data
21 predictions <- predict(rf_model, newdata = testData)
```

K-Nearest Neighbors (KNN)

To implement the K-Nearest Neighbors algorithm in R:

```
1  # Load the class package
2  library(class)
3
4  # Train the KNN model
5  knn_model <- knn(trainData[, -target_column], testData[, -target_column], trainData[, target
     _column], k = 3)
6
7  # Evaluate the model
8  accuracy <- sum(knn_model == testData[, target_column]) / nrow(testData)
```

Support Vector Machines (SVM)

To train a support vector machine model in R:

```
1  # Load the e1071 package
2  library(e1071)
3
4  # Train the SVM model
5  svm_model <- svm(target ~ ., data = trainData, kernel = "linear")
6
```

```
7 # Make predictions on test data
8 predictions <- predict(svm_model, newdata = testData)
```

Model Evaluation

To evaluate machine learning models in R:

```
1 # Load the caret package
2 library(caret)
3
4 # Evaluate the model using cross-validation
5 set.seed(123)
6 model <- train(target ~ ., data = trainData, method = "lm")
7 cv_results <- trainControl(method = "cv", number = 5)
8 cv <- train(model, trControl = cv_results)
```

Conclusion

This cheatsheet provides a comprehensive overview of R programming concepts for machine learning. It covers data manipulation, loading data, exploring data, and implementing various machine learning algorithms in R. By following this cheatsheet, beginners can get started with R for machine learning and gain a solid understanding of the fundamentals.

Author

Learn Everything AI - Shivam Modi