SathyaNarayanan Srinivasan

Sxs142031@utdallas.edu

Abstract

Construction, Pruning and Testing Decision Trees

INducing decision trees

Machine Learning Assignment - 2

Table of Contents

[Using R 2](#_Toc432891236)

[Requirements 2](#_Toc432891237)

[Prerequisites 2](#_Toc432891238)

[Packages Used 2](#_Toc432891239)

[Compiling and Running Instructions 3](#_Toc432891240)

[**Output – PLEASE NOTE (IMPORTANT)** 4](#_Toc432891241)

[Using Java 4](#_Toc432891242)

[Requirements 4](#_Toc432891243)

[Compiling and Running Instructions 4](#_Toc432891244)

[**Output** 6](#_Toc432891245)

# Using R

# Requirements

* **Operating System** – Windows 7 or Higher
* **Product Version :** [**R-3.2.2**](https://cran.r-project.org/bin/windows/base/R-3.2.2-win.exe)
* **PDF Processor** – Adobe PDF Reader

# Prerequisites

* Please ensure that R ver. 3.2.2 is installed properly and working in the machine.
* Please ensure that rscript is accessible in the command prompt.i.e. when rscript is typed in the command prompt the system recognizes that it is either a internal or an external command. If not, please follow the steps here : [Debug Link](http://stackoverflow.com/questions/17339438/r-script-from-command-line)
* This program won’t compile in RStudio as it accepts command line parameters
* Also open Command Prompt in ELEVATED PRIVILEGES mode.

# Packages Used

* Rpart
* Rpart.plot
* caTools
* caret
* partykit
* klaR
* Rgtk2
* Rattle

# Compiling and Running Instructions

* In a windows command prompt :
  + Navigate to the R folder
  + Type rscript <full path to the file Assignment2.R> with the .R extension followed by the following parameters <trainingfile> <validation file> <testFile> - Each of them is the absolute path to the file on the PC.

**EXAMPLE**

**FOR TEST SET 1**

**D:\DriveFiles\@UTD\Repositories\Assignment2> rscript assignment2.R training\_set.csv validation\_set.csv test\_set.csv**

**FOR TEST SET 2**

**D:\DriveFiles\@UTD\Repositories\Assignment2> rscript assignment2.R training\_set2.csv validation\_set2.csv test\_set2.csv**

* + Press Enter
  + Wait for the packages to be installed – it will take a while depending on many factors
  + The output is ordered according the question of assignment 2.
  + Output will be generated in the Parent Folder i.e. **~.\Assignment2\R** as **Plots.pdf**

# **Output – PLEASE NOTE (IMPORTANT)**

* **Constructed and Pruned decision trees have the same structure.**
* **“Assignment2withRattle.R” file has dependency issues with GTK+ that can be averted if run in a Mac or Windows with GTK+ already installed** - This is just to create colorful charts.

|  |  |
| --- | --- |
| **Data Set** | **Accuracy** |
| Set 1 | 73 % |
| Set 2 | 72.667% |

# Using Java

# Requirements

* **Operating System** – Windows 7 or Higher
* **Java:** [1.7.0\_45; Java HotSpot(TM) 64-Bit Server VM 24.45-b08](https://docs.oracle.com/javase/8/docs/technotes/guides/install/windows_jdk_install.html#CHDEBCCJ)
* **Runtime:** [Java(TM) SE Runtime Environment 1.7.0\_45-b18](https://docs.oracle.com/javase/8/docs/technotes/guides/install/windows_jre_install.html#CHDEDHAJ)
* **PDF Processor** – Adobe PDF Reader

# Compiling and Running Instructions

* In Elevated Windows Command Prompt:
  + Navigate to the downloaded and unzipped project folder using cd command
  + Navigate to the Java folder using cd command
  + Navigate to the folder MLID3 using cd command
  + **TO COMPILE :**

USE: **javac src/com/utd/ML/mlid3/\*.java**

* + **TO RUN**:

USE: **java -cp ./src com.utd.ML.mlid3.MLID3 <K> <L> <ABSOLUTE PATH TO TRAINING FILE> <ABSOLUTE PATH TO VALIDATION FILE> <ABSOLUTE PATH TO TESTING FILE>**

* + **Some errors as unable to find the file specified OR the inbuilt error in the program might occur. This might be due to incorrect path specification. Please give the ABSOLUTE PATH to the data sets. The folder structure might contain folders with spaces. Please enclose the entire path with “ ” in a Windows Machine. Please give the ABSOLUTE PATH correctly with respect to the OS.**
  + **Please refer to the samples below:**

**EXAMPLES:**

**COMPILING**

*D:\DriveFiles\Fall15\ML\Assignment Solutions\Assignment2\Java\MLID3> javac src/com/utd/ML/mlid3/\*.java*

**RUNNING**

**For Test Set 1**

*D:\DriveFiles\Fall15\ML\Assignment Solutions\Assignment2\Java\MLID3> java -cp ./src com.utd.ML.mlid3.MLID3 1 2 "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\training\_set1.csv" "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\validation\_set1.csv" "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\test\_set1.csv" yes*

**For Test Set 2**

*D:\DriveFiles\Fall15\ML\Assignment Solutions\Assignment2\Java\MLID3> java -cp ./src com.utd.ML.mlid3.MLID3 1 2 "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\training\_set2.csv" "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\validation\_set2.csv" "\DriveFiles\\Fall15\\ML\\Assignment Solutions\\Assignment2\\test\_set2.csv" yes*

# **Output**

|  |  |  |
| --- | --- | --- |
| **Data Set** | **Accuracy with Information Gain (%)** | **Accuracy with Variance Impurity (%)** |
| Set 1 | 76.64 % | 72.5 |
| Set 2 | 75.85% | 72.33 |