ML & DS with R (Day-1) April 11 2021

Prof. T. R. Mahore

Today's Contents

1. Machine Learning and Data Science Introduction
2. Hands on R: Programming for Machine Learning and Data Science

Machine Learning and Data Science Introduction

Course Structure

1. R - Tool
2. Data Visualisation
3. Fundamentals of Applied Stastics
4. Machine Learning Fundamentals

Machine Learning and Data Science in Scope

1. Machine Learning Fundamentals like supervised machine learnin, unsupervised machine learning etc.
2. R Programming Language - Fundamentals of R, data selection & manipulation with R, handling missiong values in R
3. Reading various files in R
4. Data visualization in R - Base functions to create bar plots, histograms, plots with libraries like ggplot2, maps, scatter3d plot & lattice
5. Applied statistics for machine for machine learning - descriptive analysis vs inferial analysis, t-test, confidence interval, standard error, standard deviation, varance, hypothesis testing and more
6. Linear Regression with R
7. Logistic Regression with R
8. Dimension Reduction techniques covering PCA
9. Clustering fundamentals with implementation of k-means in R
10. Tree based machine learning techniques like CART and Random Forest
11. KNN Implementation in R
12. Naive Bayes in R
13. Neural Network in R

Data Science Help

- To University Forum - http://www.topuniversityforum.in/forums/artificial-intelligence-machine-learning.47/

- Stack Overflow - https://stackoverflow.com/
- GOOGLE

Data Science as Career

Data Driven Managers, Digital Marketing, Machine Learning, Deep Learning, Deep Learning and Artificial Intelligence

Portfolio

a. Personal website/blog b. Github - https://github.com c. Linkedin

How to make right decision for your career in data science and machine learning

This is one of the most important decision for you and you should use following factors to determine he right choice

a. Job Profile Relevance - Exact Match, Partial Match, No Match

b. Job Experience and Current Position

c. Passion or Need

d. Other Resources like Time and Money you can spend to learn New Skills

Business Analytics

Businiss analytics is a field dedicated to make data driven decisions based on current and past data. It is closely related to business intelligence also. Both fields are essential but closely related so if you are looking for job and have experience in either of fields, you can easily cross over to others.

People in this field hardly code and they primarily use tool's like Microsoft Excel, Tableau etc. These tool's generate Reports which are used for **BA** or **BI**

Machine Learning Engineer

- You will find openings with Title **Machien Learning Engineer**
- Job responsibilities are largly technical side of **Data Science** and need expertise in **R, SQL, PYTHON** etc.

Data Scientist

- As per HBR **"Sexiest Job of 21st Century"**
- It's a broad term including all kind of roles related to data science **(AI, ML, DL)**

- Actual role and responsibilities could be far more specific like domain experts, programmer (R or Python), tool expert like Tableau or SPSS. Role is much into **Data Analysis** with technology sorrounding it

Machine Learning Fundamentals

Let's Dive

Artificial Intelligence (AI)

- It is the higher umbrella category covering all aspect of the space where machines are expected to use intelligence for decision making
- **IBM Watson** is a common example of AI Tool
- It encompasses machine learning and Deep Learning Fields
- In reality, artificial intelligence is a broad field which has been derived from Math, Computer Science, Neuroscience and Artificial Psychology

Machine Learning

- **Machine Learning** is the application of Artificial Intelligence
- Machine Learning is a Subset of AI
- Machine Learning use stastical analysis to deliver results
- In Machine Learning, you define the **Features** you need to make **Predictions** or to perform a task like E-mail Classification

Deep Learning

- Deep learning takes the automation a step ahead and you don't need to define the features
- It is practically a subset of **Machine Learning** and but different from the rest of the algorithms
- It is inspired by neuron and attempt is to make artificial neurons mimicking human neurons

ML vs DL

Machine Learning	**Deep Learning**
Need lesser data than Deep Learning	Need more data
Can work with CPU	Needs GPU for optimum performance
Need to manually define the features	System can automatically figure that out
Good & Recommended when you need to control feature defination and recreation	Recommended when your focus is on output, and not on ability to define feature but is not recommended when feature defination is important

Type of Machine Learning

1. Supervised Machine Learning

   In all types of machine learning algorithm, you train your code with existing data. If this data is Labeled and have details about the properties of data, it is called **Supervised Machine Learning**, you train your data with labeled data and then compare the results. For example, you are developing a algorithm to automatically detect Spam Emails. Non Spam Emails are called ham. So in this case, you will create a dataset with information on spam and ham emails, something like this:

Spam	**Ham**
Subject Contain Free	Subject Dosen't Contain Free
Subject Contain Win	Subject Dosen't Contain Win

   You will feed this data to your algorithm for Training Purpose, your code will learn and based on this, it will be able to make predictions in the future unlaneled dataset because it will know the rules of spam and ham emails.

2. Unsupervised Machine Learning
3. Reinforcement Machine Learning

Hand's on R: Programming for Machine Learning and Data Science

R - Introduction with Installation of R Studio

R Overview

- **R** Programming language is based on **S** language which was developed much earlier in 70's
- R Programming language was developed in 90's by **Ross Ihaka** and **Robert Gentleman** while working in the university of **Auckland**
- R is an open source **GNU** project
- Compatible with all major OS - MacOS, Linux, Unix, Windows, platforms

R Advantage

- Compatible with MacOS, Linux, Windows
- Free
- Not so steep learning curve to begin with
- Tons of packages for machine learning so our life becomes easy
- Still one of the most widely used language for machine learning

R Installation

- Step 1 - Download R - https://www.r-project.org/
- Step 2 - Download R Studio - https://www.rstudio.com/products/rstudio/download/

Vectors, Matrix and Data Frame

Know the basic before jumping into it

Vector and Matrix

- Sample *(observations)* size is 7 and there are 4 features *(vatiables)*. These 4 properties will be displayed in columns with 4 rows. First row will represent the name column.

Name	Feature1	Feature2	Feature3	Feature4
name1				
name2				
name3				
name4				
name5				
name6				
name7				

- Matrix can be represented by 7x5
- 7 Rows and 5 Columns
- Each row or column is a vector

- Entire dataset is a matrix. It is a multidimensional array (think of a spreadsheet), with multiple rows and columns
- representation: x_j^i
- i = serial/sequence number of sample
- j = serial/sequence number of dimension
- For matrix(X) - Capital letters are used, for vectors(x) - smallletters are used. But the concept remain same.
- So as in our example we have, 7 samples/observations, 4 features
- Therefore I can go upto 7 and I can go upto 4
- x_3^4: Here we are talking about 4th sample out of 7 and 3rd Feature or 3rd column in the spreadsheet. It is the index value. So if the 36rd column is "color", it means we are talking about color of 4th sample.
- Row vectors are represented by $[x_3^{49} + x_4^{50}]$
- Column vectors are represented by $[x_3^{49}]$ [+] $[x_4^{50}]$
-

Data Frame

- Tabular data structure with rows and columns
- Data frame is a stastical concept
- Usually Matrix will have only 1 type of data like numeric, character etc.
- A data frame can have multiple data types so one column could me numerical whereas other could be character

Data types in R

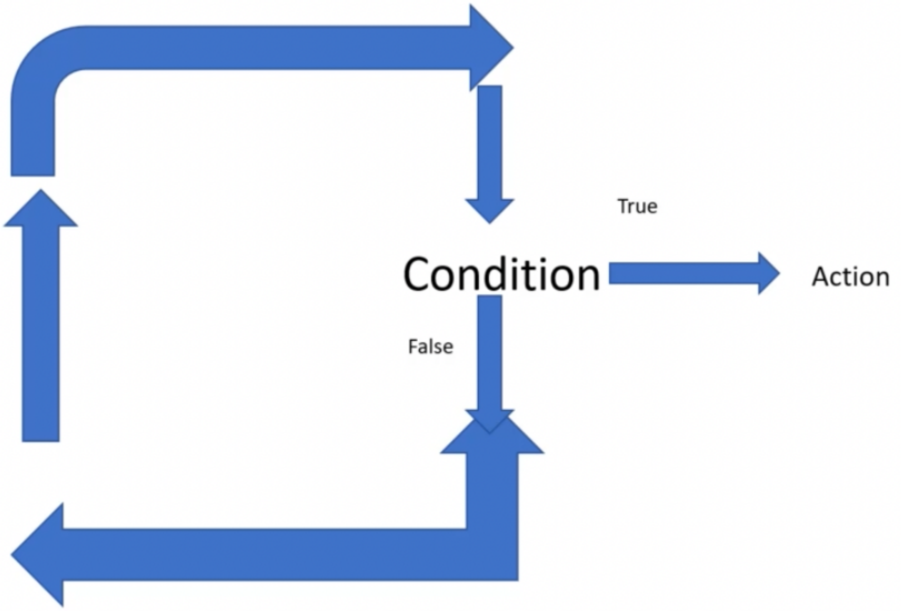
Variables & Objects

Vectors & Lists

Data Wrangling with R

Operators in R

Loops in R



If Else in R

Functions in R

Assignment

1. How do you find more information about a function in R Studio?
2. How do you install packages in R Studio?
3. You need to load a excel file in R Studio, please write a compelete command to load it. Please note you will need to do something with the library in order to load it. You can use any name for your file.
4. Load the data set and save it as a matrix.
5. Write Command to find out the class of the Data.
6. Write few data types in R.
7. Create a variable with a number.
8. Create a String Variable.
9. Print variables created in last two questions.
10. Create two numerical vectors.
11. Create an object and multiply the 2 vectors created in last question.
12. Create a list with 5 elements in R.
13. Assign names to the elements in the List.
14. Access the 1st and 4th element in the list created in question 12.
15. Remove an element from the list created in question 12.
16. Load the data frame in R Studio. After loding the data set name the columns in data set.
17. Create a new object with a column and first 5000 rows selected from data frame created in question 16.

18. Create a vector and an if-else block such that else block is executed.
19. Create a for loop to print numbers upto 5.
20. Create a while loop till number 6