

Assignment 3

Aarthi Reddy

December 1, 2017

DEADLINE: Dec 12th, 2017

1 Introduction

In this assignment you will create an R package.

2 Description of Dataset

The dataset is made available to you via elearning. The data gives the height, weight, age and gender. The training data is 'train' and the test data is 'test'. Both have 5 columns/variables:

1. Height: A numerical value
2. Weight: A numerical value
3. Age: A numerical value
4. Male: A binary value. 0 indicates female and 1 indicates male
5. Gender: Mapping to factor of 4

You will not need any other dataset.

3 Objectives

You will perform the following tasks for this assignments:

1. Write your first R package. Instructions at bottom of this assignment.
2. Write a function to plot in 3D the Age, Height and Weight of the training data color coded according to Gender
3. K Means algorithm:

- (a) Write a function that plots Age, Height and Weight in 3D color coded according to cluster, where the clusters are determined using Kmeans and K is provided by the user. In the same 3D plot mark the centroid of each cluster.
 - (b) Write a function to determine optimum number of clusters given the datasets. Explain in the comments/documentation how you arrived at that cluster number. If necessary use plots/charts etc.
 - (c) Write a function that takes in the test data, assigns a cluster to each test data point (Age, Height and Weight). Consequently determine if the data point is a male or female depending on other data points in the cluster. Provide accuracy, false positive and false negative.
4. Gaussian Mixture Models Algorithm:
- (a) Write a function that plots Age, Height and Weight in 3D color coded according to cluster, where the clusters are determined using GMM
 - (b) Write a function that assigns cluster to each test data point using information generated by mclust. Remember every test data point can belong to every cluster with a certain probability.
 - (c) Consequently determine if the data point is a male or female depending on other data points in the cluster. Provide accuracy, false positive and false negative.
5. Write a function that calls on some of the above functions and displays results obtained from KMeans and GMM. What do you think works best for this dataset?

Suggestions:

- 1. You can use mclust package for GMM. But if you find something else that works better, then use that.
- 2. Several packages can be used to do kmeans. Choose any package that works best.
- 3. You can use plotly to plot 3D. But if you find another package that works well and generates pretty plots, then please go ahead
- 4. Write clear 'Man' pages.
- 5. Imagine this package is going to be used by a business person. Your functions/man pages etc should be clear enough to be understandable by that business person
- 6. Since the kmeans and gmm are part of this one package, try and be consistent with plots and information provided. You do not want plots generated by kmeans to look very different that plots generated by GMM.

4 Expectations of student

The student is:

- Expected to complete all tasks listed above.
- Expected to submit this mini-project by 11:59pm on day of deadline mentioned at the start.
- Expected to submit original work

5 Creating an R Package

1. Install necessary packages

```
install.packages("devtools")
library("devtools")
devtools::install_github("klutometis/roxygen")
library(roxygen2)
```

2. Create your package directory

```
setwd("parent_directory")
create("cats")
```

3. Add your function; sample function:

```
cat_function <- function(love=TRUE){
  if(love==TRUE){
    print("I love cats!")
  }
  else {
    print("I am not a cool person.")
  }
}
```

Save this as a |cat_function.R|to your R directory.

4. Add comments to your R function. This is a short-cut way of maintaining documentation. Like so:

```
#' A Cat Function
#'
#' This function allows you to express your love of cats.
```

```

#' @param love Do you love cats? Defaults to TRUE.
#' @keywords cats
#' @export
#' @examples
#' cat_function()

cat_function <- function(love=TRUE){
  if(love==TRUE){
    print("I love cats!")
  }
  else {
    print("I am not a cool person.")
  }
}

```

5. Process your documentation

```

setwd("./cats")
document()

```

6. Install!

```

setwd("..")
install("cats")

```

7. Test it out. Type `?cat_function`— on your R prompt.

All sourced from <https://hilaryparker.com/2014/04/29/writing-an-r-package-from-scratch/>
 The absolute guide on building R packages is: <http://r-pkgs.had.co.nz>