Wine Quality Cluster Analysis

neelani

June 25, 2017

## Project Background

Perform a clustering analysis to understand and classify the quality of the Wines.

## Data

The data for this project is available here: Data Source: (<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>)

### Available Attributes

* 1 - fixed acidity
* 2 - volatile acidity
* 3 - citric acid
* 4 - residual sugar
* 5 - chlorides
* 6 - free sulfur dioxide
* 7 - total sulfur dioxide
* 8 - density
* 9 - pH
* 10 - sulphates
* 11 - alcohol

## Model Design Approach

Create reactive R-Shiny web application to accept the input paratmeters listed as follows, perform the k-means clustering and prodce the scatter plot

* Number of Clusters
* Attributes Selection for Clustering

### Read the data

wine <- read.csv("WineStudy/winequality-red.csv", header=TRUE, sep=";")

### Initialize the variables - Simulate the web input

* Number of Clusters - 4
* Attributes Selected - Density and Alcohol

no\_cl <- 4  
 x1 <- FALSE  
 x2 <- FALSE  
 x3 <- FALSE  
 x4 <- FALSE  
 x5 <- FALSE  
 x6 <- FALSE  
 x7 <- FALSE  
 x8 <- TRUE  
 x9 <- FALSE  
 x10 <- FALSE  
 x11 <- TRUE  
 x12 <- FALSE

### K-means clustering

colsel <- c(x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12)  
   
 wine\_sel <- wine[, colsel]  
   
 set.seed(1234)  
 cl<-kmeans(wine\_sel, no\_cl)

### Selected attributes summary

summary(wine\_sel)

## density alcohol   
## Min. :0.9901 Min. : 8.40   
## 1st Qu.:0.9956 1st Qu.: 9.50   
## Median :0.9968 Median :10.20   
## Mean :0.9967 Mean :10.42   
## 3rd Qu.:0.9978 3rd Qu.:11.10   
## Max. :1.0037 Max. :14.90

### Plot the clusters

# Plot Cluster Means   
plot(wine\_sel, col=cl$cluster)

