k‐means clustering

George Garcia

12/5/2021

WW\_test <- read.csv("~/Downloads/Website Data Sets/white\_wine\_test")  
  
WW\_training <- read.csv("~/Downloads/Website Data Sets/white\_wine\_training")

**Q11 Input and standardize both the training and test data sets.**

## Standardizing both data sets with scale() command##  
##training  
X <- subset(WW\_training, select = c("alcohol", "sugar"))  
Xs <- as.data.frame(scale(X))  
colnames(Xs) <- c("alcohol\_z", "sugar\_z")  
##test  
T <- subset(WW\_test, select = c("alcohol", "sugar"))  
Ts <- as.data.frame(scale(T))  
colnames(Ts) <- c("alcohol\_z", "sugar\_z")

**Q12 Run k‐means clustering on the training data set, using two clusters.**

#kmeans() command for 2 clusters   
kmeans01 <- kmeans(Xs, centers = 2)  
cluster <- as.factor(kmeans01$cluster)

**Q13 Give the mean of each variable within each cluster and use the means to identify a “Dry wines” and a “Sweet wines” cluster.**

#disdplay the mean of the cluster   
Cluster1 <- Xs [ which(cluster == 1),]  
Cluster2 <- Xs [ which(cluster == 2),]  
summary(Cluster1)

## alcohol\_z sugar\_z   
## Min. :-1.8265 Min. :-0.9085   
## 1st Qu.:-1.1586 1st Qu.: 0.3541   
## Median :-0.9081 Median : 0.8676   
## Mean :-0.7552 Mean : 0.9608   
## 3rd Qu.:-0.4072 3rd Qu.: 1.4882   
## Max. : 2.0138 Max. : 5.5113

summary(Cluster2)

## alcohol\_z sugar\_z   
## Min. :-1.5760 Min. :-1.1225   
## 1st Qu.:-0.1568 1st Qu.:-0.9513   
## Median : 0.4276 Median :-0.8443   
## Mean : 0.4902 Mean :-0.6236   
## 3rd Qu.: 1.1790 3rd Qu.:-0.3521   
## Max. : 2.8904 Max. : 1.4775

**Q14 Validate the clustering results by running k‐means clustering on the test data set, using two clusters, and identifying a “Dry wines” and a “Sweet wines” cluster.**

colnames(Ts) <- c("alcohol\_z", "sugar\_z")  
kmeans01\_test <- kmeans(Ts, centers = 2)  
cluster\_test <- as.factor(kmeans01\_test$cluster)  
Cluster1\_test <- Ts[ which(cluster\_test == 1), ]  
Cluster2\_test <- Ts[ which(cluster\_test == 2), ]  
summary(Cluster1\_test); summary(Cluster2\_test)

## alcohol\_z sugar\_z   
## Min. :-1.6753 Min. :-1.0891   
## 1st Qu.:-0.2187 1st Qu.:-0.9450   
## Median : 0.3680 Median :-0.8214   
## Mean : 0.4563 Mean :-0.6056   
## 3rd Qu.: 1.1570 3rd Qu.:-0.2859   
## Max. : 2.7755 Max. : 1.4235

## alcohol\_z sugar\_z   
## Min. :-2.0799 Min. :-1.0377   
## 1st Qu.:-1.1897 1st Qu.: 0.3963   
## Median :-0.9470 Median : 1.0322   
## Mean :-0.8024 Mean : 1.0650   
## 3rd Qu.:-0.5424 3rd Qu.: 1.5832   
## Max. : 1.5616 Max. : 3.2978