1.

#isye 6501 hw2 4.2 pamk method

library(cluster)

library(NbClust)

library(fpc)

library(vegan)

#input data

data <- read.table('C:\\Users\\huangchengqi\\Desktop\\MS SCE\\19Fall\\ISYE6501\\hw2\\iris.txt',header=TRUE)

raw\_data <- as.matrix(data[,-5])

data2 <- scale(raw\_data, center=TRUE, scale=TRUE)

#pamk funcion

pamk.best1=pamk(data2[,c(1,2)])$nc

pamk.best2=pamk(data2[,c(1,3)])$nc

pamk.best3=pamk(data2[,c(1,4)])$nc

pamk.best4=pamk(data2[,c(2,3)])$nc

pamk.best5=pamk(data2[,c(2,4)])$nc

pamk.best6=pamk(data2[,c(3,4)])$nc

pamk.best7=pamk(data2[,c(1,2,3)])$nc

pamk.best8=pamk(data2[,c(1,2,4)])$nc

pamk.best9=pamk(data2[,c(1,3,4)])$nc

pamk.best10=pamk(data2[,c(2,3,4)])$nc

pamk.best11=pamk(data2[,c(1,2,3,4)])$nc

k\_result <- matrix(c(pamk.best1,pamk.best2,pamk.best3,pamk.best4,pamk.best5,pamk.best6,pamk.best7,pamk.best8,pamk.best9,pamk.best10,pamk.best11),nrow=1,ncol=11)

colnames(k\_result) <- c('sl+sw','sl+pl','sl+pw','sw+pl','sw+pw',

'pl+pw','sl+sw+pl','sl+sw+pw','sl+pl+pw',

'sw+pl+pw','sl+sw+pl+pw')

2.

#isye 6501 hw2 4.2 elbow method

library(cluster)

library(NbClust)

library(fpc)

library(vegan)

#input data

data <- read.table('C:\\Users\\huangchengqi\\Desktop\\MS SCE\\19Fall\\ISYE6501\\hw2\\iris.txt',header=TRUE)

raw\_data <- as.matrix(data[,-5])

data2 <- scale(raw\_data, center=TRUE, scale=TRUE)

#elbow method, using sum of squared error

wssplot <- function(data,nc=10,seed=1234){

wss <- (nrow(data)-1)\*sum(apply(data,2,var))

for (i in 2:nc){

set.seed(seed)

wss[i] <- sum(kmeans(data,centers=i)$withinss)

}

plot(1:nc,wss,type='b',xlab='number of clusters',

ylab='Within groups sum of squares' )

}

wssplot(data2)

wssplot\_2 <- function(data,color\_type){

wss <- (nrow(data)-1)\*sum(apply(data,2,var))

for (i in 2:10){

set.seed(1234)

wss[i] <- sum(kmeans(data,centers=i)$withinss)

}

lines(1:10,wss,type='b',col=color\_type,xlab='number of clusters',

ylab='Within groups sum of squares' )

}

wssplot\_2(data2[,c(1,2)],2)

wssplot\_2(data2[,c(1,3)],3)

wssplot\_2(data2[,c(1,4)],4)

wssplot\_2(data2[,c(2,3)],5)

wssplot\_2(data2[,c(2,4)],6)

wssplot\_2(data2[,c(3,4)],7)

wssplot\_2(data2[,c(1,2,3)],8)

wssplot\_2(data2[,c(2,3,4)],9)

wssplot\_2(data2[,c(1,2,4)],10)

wssplot\_2(data2[,c(1,3,4)],11)

#k=2

model <- kmeans(data2,2)

pred <- model$cluster

result\_matrix <- matrix(c(data[,5],pred),nrow=150,ncol=2)

#k=3

model2 <- kmeans(data2,3)

pred2 <- model2$cluster

result\_matrix2 <- matrix(c(data[,5],pred2),nrow=150,ncol=2)

#pam,k=2

km\_stats1 <- cluster.stats(dist(data2), model$cluster)

ave\_sil1=km\_stats1$avg.silwidth

ave\_with1=km\_stats1$average.within

ave\_bet1=km\_stats1$average.between

#pam,k=3

km\_stats2 <- cluster.stats(dist(data2), model2$cluster)

ave\_sil2=km\_stats2$avg.silwidth

ave\_with2=km\_stats2$average.within

ave\_bet2=km\_stats2$average.between

result\_table <- matrix(nrow=2,ncol=3)

colnames(result\_table) <- c('Silwidth','AverageWithin','averageBetween')

result\_table[1,1]=ave\_sil1

result\_table[1,2]=ave\_with1

result\_table[1,3]=ave\_bet1

result\_table[2,1]=ave\_sil2

result\_table[2,2]=ave\_with2

result\_table[2,3]=ave\_bet2