Mingjie Zhao

a

wcgs = read.table("wcgs2.dat",header=T)

# K-means with K = 3

km1 = kmeans(wcgs[,1:4],3)

# K-means with K = 2

km2 = kmeans(wcgs[,1:4],2)

# Associate with high-grade status

# K = 3 and chd

chisq.test(table(km1$cluster,wcgs[,5]))

Pearson's Chi-squared test

data: table(km1$cluster, wcgs[, 5])

X-squared = 23.8131, df = 2, p-value = 6.746e-06

b

# K = 2 and chd

chisq.test(table(km2$cluster,wcgs[,5]))

Pearson's Chi-squared test with Yates' continuity correction

data: table(km2$cluster, wcgs[, 5])

X-squared = 18.0647, df = 1, p-value = 2.135e-05

c

# K = 3 and behavior

chisq.test(table(km1$cluster,wcgs[,6]))

Pearson's Chi-squared test

data: table(km1$cluster, wcgs[, 6])

X-squared = 13.7684, df = 2, p-value = 0.001024

d

# K = 2 and behavior

chisq.test(table(km2$cluster,wcgs[,6]))

Pearson's Chi-squared test with Yates' continuity correction

data: table(km2$cluster, wcgs[, 6])

X-squared = 5.7908, df = 1, p-value = 0.01611

e

std=scale(wcgs[,1:4],center=TRUE,scale=TRUE)

# K-means with K = 2

km2 = kmeans(std[,1:4],2)

#K = 2 and chd

chisq.test(table(km2$cluster,wcgs[,5]))

Pearson's Chi-squared test with Yates' continuity correction

data: table(km2$cluster, wcgs[, 5])

X-squared = 9.8608, df = 1, p-value = 0.001688

f

#K = 2 and behavior

chisq.test(table(km2$cluster,wcgs[,6]))

Pearson's Chi-squared test with Yates' continuity correction

data: table(km2$cluster, wcgs[, 6])

X-squared = 3.7047, df = 1, p-value = 0.05426

g

bickm = function(km)

{b = length(km$centers[1,])

sum = sum(km$size)

k = length(unique(km$cluster))

p = k\*(b+1)

bic = sum(km$withinss)+p\*log(sum)/2

bic

}

bic=NULL

for(i in 1:10){

temp=kmeans(wcgs[,1:4],i)

temp.bic=bickm(temp)

bic=c(bic,temp.bic)

}

bic

[1] 2240183.4 1346408.8 1081920.8 823097.4 728925.2 628963.4 563897.1

[8] 523453.5 487520.4 450847.7

The model when K=10 gives the smallest BIC value.

h

(c):

H\_0: when k=3, Cluster and Behavior Type are independent

(e):

H\_0: when k=2, Cluster based on standardized data and Behavior Type are independent