Tianqian Tan

Classification HW

# Use setwd() to point to the directory containing the files

# or you may use the choose.dir() function which will give a pop up window

c1.df = read.table( file = 'class.1.dat') # read in data frame

str(c1.df) # should be 3000 observations

c1 = c1.df$x

c2.df = read.table( file = 'class.2.dat') # read in data frame

str(c2.df) # should be 3000 observations

c2 = c2.df$x

new.obs.df = read.table( file = 'new.obs.dat') # read in data frame

str(new.obs.df) # should be 1000 observations

new.obs = new.obs.df$x

class.assignment = numeric(1000)

###### Now write code to classify the new observations ###############

## if a point is in C1 we declare it is from the first population

## if a point is in C2 we declare it is from the second population

## plot first in red and second in blue and new obs in Orange

x.hw<-rbind(c1,c2)

plot(x.hw, type = 'n', main = "Equal Cost and a priori Prob")

points(c1,col = 'red')

points(c2,col = 'blue')

points(new.obs,col="Orange")

#density estimator

c1.pdf=density(c1,bw=9)

plot(c1.pdf$x,c1.pdf$y,type="l") #plot looks good

c2.pdf=density(c2,bw=10)

plot(c2.pdf$x,c2.pdf$y,type="l") #plot looks good

pdf.lookup.c1=pdf.hat(c1,bw=9,n=3000)

pdf.lookup.c2=pdf.hat(c2,bw=10,n=3000)

plot(pdf.lookup.c1$pdf,type='l')

points(pdf.lookup.c2$pdf,col="red",type='l')

pdf.vector.c1=pdf.lookup(pdf.lookup.c1,new.obs)

pdf.vector.c2=pdf.lookup(pdf.lookup.c2,new.obs)

class=pdf.vector.c1-pdf.vector.c2

#classify data

class.assignment=numeric(length(class))

for (i in 1: length(class))

{ if(class[i]>0)

{ class.assignment[i]=1

}

else

{ class.assignment[i]=2

}

}

class.assignment

table(class.assignment)

write.table(class.assignment, file = 'Tianqian.Tan.classes.dat')