Assignment-5

Ishaan Yadav Cs-D 74

#### Consider 18 points data set referred in theory class.

df <- read.csv("/Users/Rajnandini/Desktop/rstudio/Expt 5-Data Set\_knn1\_csv.csv")

#### Consider test sample P(3,2). Apply following algorithms to find the class of this test point.

#### i. NN

##### Create a distance column to store distances of all training points from the test point P(3, 2)

df$distance<-sqrt((3-df$x)^2+(2-df$y)^2)  
head(df)

## sno x y class distance  
## 1 1 0.8 0.8 1 2.505993  
## 2 2 1.0 1.0 1 2.236068  
## 3 3 1.2 0.8 1 2.163331  
## 4 4 0.8 1.2 1 2.340940  
## 5 5 1.2 1.2 1 1.969772  
## 6 6 4.0 3.0 2 1.414214

##### Sort the data frame in ascending order of distance column

sorted\_df <- df[order(df$distance),]  
head(sorted\_df)

## sno x y class distance  
## 16 16 3.5 1.0 3 1.118034  
## 7 7 3.8 2.8 2 1.131371  
## 14 14 3.2 0.7 3 1.315295  
## 6 6 4.0 3.0 2 1.414214  
## 17 17 4.0 1.0 3 1.414214  
## 8 8 4.2 2.8 2 1.442221

##### Prediction using NN:

prediction\_NN<-sorted\_df[1,4]  
cat("Predicted class for P using NN is: ", prediction\_NN)

## Predicted class for P using NN is: 3

#### ii. KNN with K=5 and K=7

##### Extract 5 nearest neighbours

NN5<-sorted\_df[1:5,]  
NN5

## sno x y class distance  
## 16 16 3.5 1.0 3 1.118034  
## 7 7 3.8 2.8 2 1.131371  
## 14 14 3.2 0.7 3 1.315295  
## 6 6 4.0 3.0 2 1.414214  
## 17 17 4.0 1.0 3 1.414214

##### Prediction using 5 neighbours

#table(NN5$class)  
prediction\_NN5<-names(which.max(table(NN5$class)))  
cat("Predicted class using 5 nearest neighbours is: ", prediction\_NN5)

## Predicted class using 5 nearest neighbours is: 3

##### Extract 7 nearest neighbours

NN7<-sorted\_df[1:7,]  
NN7

## sno x y class distance  
## 16 16 3.5 1.0 3 1.118034  
## 7 7 3.8 2.8 2 1.131371  
## 14 14 3.2 0.7 3 1.315295  
## 6 6 4.0 3.0 2 1.414214  
## 17 17 4.0 1.0 3 1.414214  
## 8 8 4.2 2.8 2 1.442221  
## 9 9 3.8 3.2 2 1.442221

##### Prediction using 7 neighbours

#table(NN5$class)  
prediction\_NN7<-names(which.max(table(NN7$class)))  
cat("Predicted class using 7 nearest neighbours is: ", prediction\_NN7)

## Predicted class using 7 nearest neighbours is: 2

#### iii. MKNN with K=5

##### Create a weight column in NN5 dataframe to store weight of each of the 5 neighbours

NN5$weight<-(NN5$distance[5]-NN5$distance)/(NN5$distance[5]-NN5$distance[1])  
NN5

## sno x y class distance weight  
## 16 16 3.5 1.0 3 1.118034 1.0000000  
## 7 7 3.8 2.8 2 1.131371 0.9549704  
## 14 14 3.2 0.7 3 1.315295 0.3339829  
## 6 6 4.0 3.0 2 1.414214 0.0000000  
## 17 17 4.0 1.0 3 1.414214 0.0000000

##### Create a grouped table to find the classwise sum of weights

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

class\_sum<-NN5%>%group\_by(class)%>%summarize(sum\_of\_weights=sum(weight))  
class\_sum

## # A tibble: 2 × 2  
## class sum\_of\_weights  
## <int> <dbl>  
## 1 2 0.955  
## 2 3 1.33

[Note: %>% -a type operator, used for nesting ]

##### Prediction using MKNN

prediction\_MKNN<-class\_sum$class[which.max(class\_sum$sum\_of\_weights)]  
cat("Predicted class using MKNN is: ", prediction\_MKNN)

## Predicted class using MKNN is: 3

#### iv) R-NN Radius based algorithm with radius as 1.45 units.

##### Create R- neighbourhood dataframe containing data points whose distance from P is less than 1.45

r\_neighbours <- df[df$distance<1.45,]  
r\_neighbours

## sno x y class distance  
## 6 6 4.0 3.0 2 1.414214  
## 7 7 3.8 2.8 2 1.131371  
## 8 8 4.2 2.8 2 1.442221  
## 9 9 3.8 3.2 2 1.442221  
## 14 14 3.2 0.7 3 1.315295  
## 16 16 3.5 1.0 3 1.118034  
## 17 17 4.0 1.0 3 1.414214

##### Prediction using R neighbours

prediction\_RNN <- names(which.max(table(r\_neighbours$class)))  
cat("Predicted class using R neighbours is ", prediction\_RNN)

## Predicted class using R neighbours is 2