Poorvi_Raut_HWO3_KNN.R

Owner

2023-02-27

```
#knowledge Discovery and Data Mining (CS 513) Homework 3: KNN classification
#Course : CS 513-A
# First Name : Poorvi
#Last Name : Raut
# ID : 20009560
# Purpose : HW_03_KNN

#clearing object environment
rm(list = ls())

#get working directory
getwd()
```

```
## [1] "C:/Users/Owner/Desktop/Spring 2023/CS 513 KDD"
```

```
#Load the "breast-cancer-wisconsin.data.csv" from canvas into R and perform the KNN classificati
on
dataSet1<-read.csv("/Users/Owner/Desktop/Spring 2023/CS 513 KDD/breast-cancer-wisconsin.csv",na.
string = "?" )
View(dataSet1)

n<-as.numeric(as.character(dataSet1$F6))
dataSet1$F6<-n

#Remove the rows with missing values
dataSet1<-na.omit(dataSet1)

#Converting labels to factor class
dataSet1$Class<-factor(dataSet1$Class,levels = c("2","4"),labels = c("benign","malignant"))

#knn classification
#Generating  training set and testing set in the ratio 70% to 30%
size<-sample(1:nrow(dataSet1),0.7*nrow(dataSet1))
n1<-function(x){(x-min(x))/max(x)-min(x)}

#Running min-max normalization on first 4 columns since they are predictor attributes
norm<-as.data.frame(lapply(dataSet1[,c(2,3,4,5,6,7,8,9,10)],n1))

dataSet2<-dataSet1['Class']
#training set
train<-norm[size,]
train2<-dataSet2[size,]

#testing set
test<-norm[size,]
test2<-dataSet2[size,]

#Loading the package class
library(class)

#running  KNN function for k = 3
#classifier<-knn(train[, -1], test[, -1],cl=train2,k=3)
#clf <- knn(train,test,cl=train2,k=3)
clf <- knn(train,test,cl=train2,k=3)
#creating confusion matrix
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
conf_matrix<-table(clf,test2)
print(conf_matrix)
```

```
##           test2
## clf         benign malignant
##   benign       314         2
##   malignant      8       154
```

```
confusionMatrix(clf, test2)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign       314         2
##   malignant      8       154
##
##                Accuracy : 0.9791
##                  95% CI : (0.9619, 0.9899)
##     No Information Rate : 0.6736
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9529
##
##  Mcnemar's Test P-Value : 0.1138
##
##             Sensitivity : 0.9752
##             Specificity : 0.9872
##          Pos Pred Value : 0.9937
##          Neg Pred Value : 0.9506
##              Prevalence : 0.6736
##          Detection Rate : 0.6569
##    Detection Prevalence : 0.6611
##       Balanced Accuracy : 0.9812
##
##        'Positive' Class : benign
##
```

```
#Accuracy calculation
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix)
```

```
## [1] 97.90795
```

```
#running  KNN function for k = 5
clf <- knn(train,test,cl=train2,k=5)
#creating confusion matrix
library(caret)
conf_matrix1<-table(clf,test2)
print(conf_matrix1)
```

```
##           test2
## clf         benign malignant
##   benign       317         5
##   malignant      5       151
```

```
confusionMatrix(clf, test2)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign       317         5
##   malignant      5       151
##
##                Accuracy : 0.9791
##                  95% CI : (0.9619, 0.9899)
##     No Information Rate : 0.6736
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9524
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9845
##             Specificity : 0.9679
##          Pos Pred Value : 0.9845
##          Neg Pred Value : 0.9679
##              Prevalence : 0.6736
##          Detection Rate : 0.6632
##    Detection Prevalence : 0.6736
##       Balanced Accuracy : 0.9762
##
##        'Positive' Class : benign
##
```

```
#Accuracy calculation
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix1)
```

```
## [1] 97.90795
```

```
#running  KNN function for k = 10
clf <- knn(train,test,cl=train2,k=10)
#creating confusion matrix
library(caret)
conf_matrix2<-table(clf,test2)
print(conf_matrix2)
```

```
##           test2
## clf         benign malignant
##   benign       315         8
##   malignant      7       148
```

```
confusionMatrix(clf, test2)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign       315         8
##   malignant      7       148
##
##                Accuracy : 0.9686
##                  95% CI : (0.9488, 0.9823)
##     No Information Rate : 0.6736
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9285
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9783
##             Specificity : 0.9487
##          Pos Pred Value : 0.9752
##          Neg Pred Value : 0.9548
##              Prevalence : 0.6736
##          Detection Rate : 0.6590
##    Detection Prevalence : 0.6757
##       Balanced Accuracy : 0.9635
##
##        'Positive' Class : benign
##
```

```
#Accuracy calculation
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix2)
```

```
## [1] 96.86192
```

```
#running  KNN function for k = 15
clf <- knn(train,test,cl=train2,k=15)
#creating confusion matrix
library(caret)
conf_matrix2<-table(clf,test2)
print(conf_matrix2)
```

```
##           test2
## clf         benign malignant
##   benign       316        10
##   malignant      6       146
```

```
confusionMatrix(clf, test2)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign       316        10
##   malignant      6       146
##
##                Accuracy : 0.9665
##                  95% CI : (0.9462, 0.9807)
##     No Information Rate : 0.6736
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9234
##
##  Mcnemar's Test P-Value : 0.4533
##
##             Sensitivity : 0.9814
##             Specificity : 0.9359
##          Pos Pred Value : 0.9693
##          Neg Pred Value : 0.9605
##              Prevalence : 0.6736
##          Detection Rate : 0.6611
##    Detection Prevalence : 0.6820
##       Balanced Accuracy : 0.9586
##
##        'Positive' Class : benign
##
```

```
#Accuracy calculation
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix2)
```

```
## [1] 96.65272
```

```
#running  KNN function for k = 25
clf <- knn(train,test,cl=train2,k=25)
#creating confusion matrix
library(caret)
conf_matrix2<-table(clf,test2)
print(conf_matrix2)
```

```
##           test2
## clf         benign malignant
##   benign       316        12
##   malignant      6       144
```

```
confusionMatrix(clf, test2)
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  benign malignant
##   benign       316        12
##   malignant      6       144
##
##                Accuracy : 0.9623
##                  95% CI : (0.9411, 0.9775)
##     No Information Rate : 0.6736
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9135
##
##  Mcnemar's Test P-Value : 0.2386
##
##             Sensitivity : 0.9814
##             Specificity : 0.9231
##          Pos Pred Value : 0.9634
##          Neg Pred Value : 0.9600
##              Prevalence : 0.6736
##          Detection Rate : 0.6611
##    Detection Prevalence : 0.6862
##       Balanced Accuracy : 0.9522
##
##        'Positive' Class : benign
##
```

```
#Accuracy calculation
accuracy<-function(x){sum(diag(x)/sum(rowSums(x)))*100}
accuracy(conf_matrix2)
```

```
## [1] 96.23431
```

```
#### END OF ASSIGNMENT ###
```