HarshAgrawal_HW9.R

Asus

2021-12-03

```
# Course      : CS 513
# First Name  : Harsh
# Last Name   : Agrawal
# CWID        : 10475285

rm(list=ls())
setwd("C:/Users/Asus/Desktop/Github/CS513_KnowledgeDiscoveryDataMining/HW9")

# Reading and removing NAs and iD column
db = read.csv('wisc_bc_ContinuousVar.csv',header=TRUE, sep=",")
db = db[,-1]
db<-na.omit(db)
summary(db)

##   diagnosis          radius_mean      texture_mean   perimeter_mean
##  Length:569         Min.   : 6.981   Min.   : 9.71   Min.   : 43.79
##  Class :character   1st Qu.:11.700   1st Qu.:16.17   1st Qu.: 75.17
##  Mode  :character   Median :13.370   Median :18.84   Median : 86.24
##                     Mean   :14.127   Mean   :19.29   Mean   : 91.97
##                     3rd Qu.:15.780   3rd Qu.:21.80   3rd Qu.:104.10
##                     Max.   :28.110   Max.   :39.28   Max.   :188.50
##    area_mean      smoothness_mean   compactness_mean  concavity_mean
##  Min.   : 143.5   Min.   :0.05263   Min.   :0.01938   Min.   :0.00000
##  1st Qu.: 420.3   1st Qu.:0.08637   1st Qu.:0.06492   1st Qu.:0.02956
##  Median : 551.1   Median :0.09587   Median :0.09263   Median :0.06154
##  Mean   : 654.9   Mean   :0.09636   Mean   :0.10434   Mean   :0.08880
##  3rd Qu.: 782.7   3rd Qu.:0.10530   3rd Qu.:0.13040   3rd Qu.:0.13070
##  Max.   :2501.0   Max.   :0.16340   Max.   :0.34540   Max.   :0.42680
##  concave.points_mean symmetry_mean    fractal_dimension_mean   radius_se
##  Min.   :0.00000     Min.   :0.1060   Min.   :0.04996        Min.
:0.1115
##  1st Qu.:0.02031     1st Qu.:0.1619   1st Qu.:0.05770        1st
Qu.:0.2324
##  Median :0.03350     Median :0.1792   Median :0.06154        Median
:0.3242
##  Mean   :0.04892     Mean   :0.1812   Mean   :0.06280        Mean
:0.4052
##  3rd Qu.:0.07400     3rd Qu.:0.1957   3rd Qu.:0.06612        3rd
Qu.:0.4789
##  Max.   :0.20120     Max.   :0.3040   Max.   :0.09744        Max.
:2.8730
##    texture_se      perimeter_se       area_se        smoothness_se
```

```
##  Min.   :0.3602   Min.   : 0.757   Min.   :  6.802   Min.   :0.001713
##  1st Qu.:0.8339   1st Qu.: 1.606   1st Qu.: 17.850   1st Qu.:0.005169
##  Median :1.1080   Median : 2.287   Median : 24.530   Median :0.006380
##  Mean   :1.2169   Mean   : 2.866   Mean   : 40.337   Mean   :0.007041
##  3rd Qu.:1.4740   3rd Qu.: 3.357   3rd Qu.: 45.190   3rd Qu.:0.008146
##  Max.   :4.8850   Max.   :21.980   Max.   :542.200   Max.   :0.031130
##  compactness_se      concavity_se     concave.points_se   symmetry_se
##  Min.   :0.002252   Min.   :0.00000   Min.   :0.000000   Min.   :0.007882
##  1st Qu.:0.013080   1st Qu.:0.01509   1st Qu.:0.007638   1st Qu.:0.015160
##  Median :0.020450   Median :0.02589   Median :0.010930   Median :0.018730
##  Mean   :0.025478   Mean   :0.03189   Mean   :0.011796   Mean   :0.020542
##  3rd Qu.:0.032450   3rd Qu.:0.04205   3rd Qu.:0.014710   3rd Qu.:0.023480
##  Max.   :0.135400   Max.   :0.39600   Max.   :0.052790   Max.   :0.078950
##  fractal_dimension_se  radius_worst   texture_worst   perimeter_worst
##  Min.   :0.0008948    Min.   : 7.93   Min.   :12.02   Min.   : 50.41
##  1st Qu.:0.0022480    1st Qu.:13.01   1st Qu.:21.08   1st Qu.: 84.11
##  Median :0.0031870    Median :14.97   Median :25.41   Median : 97.66
##  Mean   :0.0037949    Mean   :16.27   Mean   :25.68   Mean   :107.26
##  3rd Qu.:0.0045580    3rd Qu.:18.79   3rd Qu.:29.72   3rd Qu.:125.40
##  Max.   :0.0298400    Max.   :36.04   Max.   :49.54   Max.   :251.20
##    area_worst     smoothness_worst  compactness_worst concavity_worst
##  Min.   : 185.2   Min.   :0.07117   Min.   :0.02729   Min.   :0.0000
##  1st Qu.: 515.3   1st Qu.:0.11660   1st Qu.:0.14720   1st Qu.:0.1145
##  Median : 686.5   Median :0.13130   Median :0.21190   Median :0.2267
##  Mean   : 880.6   Mean   :0.13237   Mean   :0.25427   Mean   :0.2722
##  3rd Qu.:1084.0   3rd Qu.:0.14600   3rd Qu.:0.33910   3rd Qu.:0.3829
##  Max.   :4254.0   Max.   :0.22260   Max.   :1.05800   Max.   :1.2520
##  concave.points_worst symmetry_worst   fractal_dimension_worst
##  Min.   :0.00000      Min.   :0.1565   Min.   :0.05504
##  1st Qu.:0.06493      1st Qu.:0.2504   1st Qu.:0.07146
##  Median :0.09993      Median :0.2822   Median :0.08004
##  Mean   :0.11461      Mean   :0.2901   Mean   :0.08395
##  3rd Qu.:0.16140      3rd Qu.:0.3179   3rd Qu.:0.09208
##  Max.   :0.29100      Max.   :0.6638   Max.   :0.20750
```

```r
#Splitting into training and testing
train_index <- sample(nrow(db),as.integer(.70*nrow(db)))
train_data<-db[train_index,]
test_data<-db[-train_index,]

#Factorizing
train_data$diagnosis <- factor(train_data$diagnosis)
test_data$diagnosis <- factor(test_data$diagnosis)

#SVM model building
library(e1071)
svm.model <- svm( diagnosis~ ., data = train_data  )
svm.pred <- predict(svm.model, test_data )
```

```r
#Testing accuracy and printing model parameters
table(actual=test_data[,1],svm.pred )
```

```
##       svm.pred
## actual   B   M
##      B 111   1
##      M   1  58
```

```r
SVM_wrong<- (test_data$diagnosis!=svm.pred)
rate<-sum(SVM_wrong)/length(SVM_wrong)
accuracy = 1-rate
accuracy
```

```
## [1] 0.9883041
```

```r
print(svm.model)
```

```
##
## Call:
## svm(formula = diagnosis ~ ., data = train_data)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  radial
##        cost:  1
##
## Number of Support Vectors:  104
```