

Implement SVM classification

Harini G

```
data = read.csv("D:/Harini(christ unniversity)/2nd sem subjects/R/heart  
disease.csv")
```

```
head(data)
```

```
##   age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 1  63  1  3    145   233   1        0    150    0     2.3    0  0    1
## 2  37  1  2    130   250   0        1    187    0     3.5    0  0    2
## 3  41  0  1    130   204   0        0    172    0     1.4    2  0    2
## 4  56  1  1    120   236   0        1    178    0     0.8    2  0    2
## 5  57  0  0    120   354   0        1    163    1     0.6    2  0    2
## 6  57  1  0    140   192   0        1    148    0     0.4    1  0    1
##   target
## 1      1
## 2      1
## 3      1
## 4      1
## 5      1
## 6      1
```

```
summary(data)
```

```
##           age           sex           cp           trestbps
##  Min.      :29.00   Min.      :0.0000   Min.      :0.000   Min.      : 94.0
## 1st Qu.:47.50   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:120.0
##  Median :55.00   Median :1.0000   Median :1.000   Median :130.0
##  Mean     :54.37   Mean     :0.6832   Mean     :0.967   Mean     :131.6
## 3rd Qu.:61.00   3rd Qu.:1.0000   3rd Qu.:2.000   3rd Qu.:140.0
##  Max.     :77.00   Max.     :1.0000   Max.     :3.000   Max.     :200.0
##           chol           fbs           restecg           thalach
##  Min.     :126.0   Min.     :0.0000   Min.     :0.0000   Min.     : 71.0
## 1st Qu.:211.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:133.5
##  Median :240.0   Median :0.0000   Median :1.0000   Median :153.0
##  Mean     :246.3   Mean     :0.1485   Mean     :0.5281   Mean     :149.6
## 3rd Qu.:274.5   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:166.0
##  Max.     :564.0   Max.     :1.0000   Max.     :2.0000   Max.     :202.0
##           exang           oldpeak           slope           ca
##  Min.     :0.0000   Min.     :0.00   Min.     :0.000   Min.     :0.0000
## 1st Qu.:0.0000   1st Qu.:0.00   1st Qu.:1.000   1st Qu.:0.0000
##  Median :0.0000   Median :0.80   Median :1.000   Median :0.0000
##  Mean     :0.3267   Mean     :1.04   Mean     :1.399   Mean     :0.7294
## 3rd Qu.:1.0000   3rd Qu.:1.60   3rd Qu.:2.000   3rd Qu.:1.0000
##  Max.     :1.0000   Max.     :6.20   Max.     :2.000   Max.     :4.0000
##           thal           target
##  Min.     :0.000   Min.     :0.0000
## 1st Qu.:2.000   1st Qu.:0.0000
##  Median :2.000   Median :1.0000
##  Mean     :2.314   Mean     :0.5446
```

```

## 3rd Qu.:3.000 3rd Qu.:1.0000
## Max. :3.000 Max. :1.0000

sapply(data, class)#displaying the datatype of each column

##      age      sex      cp trestbps      chol      fbs  restecg
thalach
## "integer" "integer" "integer" "integer" "integer" "integer" "integer"
"integer"
##      exang  oldpeak      slope      ca      thal      target
## "integer" "numeric" "integer" "integer" "integer" "integer"

library(caTools)

## Warning: package 'caTools' was built under R version 4.0.4

colSums(is.na(data))#checking if there are any null values

##      age      sex      cp trestbps      chol      fbs  restecg  thalach
##      0      0      0      0      0      0      0      0
##      exang  oldpeak      slope      ca      thal      target
##      0      0      0      0      0      0      0

#splitting the dataset into training and testing
sample=sample.split(data$target,SplitRatio=0.75)
train=subset(data,sample==TRUE)
test=subset(data,sample==FALSE)
dim(train)#dimesion of train data

## [1] 228 14

dim(test)

## [1] 75 14

library(e1071)

classifier = svm(formula = target ~ .,
                 data = train,
                 type = 'C-classification',
                 kernel = 'linear')

# Predicting the Test set results
y_pred = predict(classifier, newdata = test)
y_pred

## 1 2 6 7 12 13 15 16 23 24 29 30 42 44 45 49 52 56
57 67
## 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1

```

```
## 68 79 87 90 93 94 106 109 110 114 115 117 118 119 121 127 141 143
153 155
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1
1 1
## 161 178 184 193 195 198 199 200 202 203 207 210 211 214 218 222 225 226
231 233
## 1 1 0 0 1 0 0 1 0 0 0 0 1 0 0 0 0 0
1 0
## 244 249 250 261 267 268 271 272 273 278 281 282 286 292 300
## 0 0 0 0 0 1 0 1 0 1 0 1 0 0 1
## Levels: 0 1
```

```
summary(y_pred)
```

```
## 0 1
## 26 49
```

```
library(caret)
```

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

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

```
confusionMatrix(table(y_pred, test$target))
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##
```

```
## y_pred 0 1
```

```
##      0 24 2
```

```
##      1 10 39
```

```
##
```

```
##              Accuracy : 0.84
```

```
##              95% CI : (0.7372, 0.9145)
```

```
##      No Information Rate : 0.5467
```

```
##      P-Value [Acc > NIR] : 7.56e-08
```

```
##
```

```
##              Kappa : 0.6706
```

```
##
```

```
## Mcnemar's Test P-Value : 0.04331
```

```
##
```

```
##              Sensitivity : 0.7059
```

```
##              Specificity : 0.9512
```

```
##              Pos Pred Value : 0.9231
```

```
##              Neg Pred Value : 0.7959
```

```
##              Prevalence : 0.4533
```

```
##              Detection Rate : 0.3200
```

```
##      Detection Prevalence : 0.3467
```

```
##      Balanced Accuracy : 0.8286
```

```
##
```

```
##      'Positive' Class : 0
##
```