CS 513 HW7 - SVM

Aidan Fischer - 10447681

2023-11-27

I plege my honor that I have abided by the Stevens Honor System

Creation (Copied from HW3 since the data setup is the same)

```
rm(list=ls())
library(caTools)
library(class)
library(e1071)
library(caret)
```

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

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

```
## Warning: package 'lattice' was built under R version 4.3.2
```

```
library(C50)
```

```
## Warning: package 'C50' was built under R version 4.3.2
```

```
data = read.csv("breast-cancer-wisconsin.csv")
data$F6 <- suppressWarnings(as.numeric(data$F6))
data = data[complete.cases(data), ]
#Convert categories to the factor data type

for(i in 1:9){
  col = paste("F",i,sep='')
  data[col] <- factor(data[[col]], levels = 1:10)
}

data$Class <- factor(data$Class, levels=c(2,4))
set.seed(255)
split = sample.split(data$Class, SplitRatio=0.7)
train = subset(data, split == TRUE)
test = subset(data, split == FALSE)
```

SVM Model (linear kernel)

```
classifier = svm(Class ~ ., data=data, kernel = "linear")
```

Evaluation

```
train_pred <- predict(classifier, newdata=train, type="class")
test_pred  <- predict(classifier, newdata=test, type="class")

cm_train <- table(train$Class, train_pred)
cm_test <- table(test$Class, test_pred)
```

```r
confusionMatrix(cm_train)
```

```
## Confusion Matrix and Statistics
##
##    train_pred
##       2   4
##   2 308   3
##   4   2 165
##
##                Accuracy : 0.9895
##                  95% CI : (0.9758, 0.9966)
##     No Information Rate : 0.6485
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.977
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9935
##             Specificity : 0.9821
##          Pos Pred Value : 0.9904
##          Neg Pred Value : 0.9880
##              Prevalence : 0.6485
##          Detection Rate : 0.6444
##    Detection Prevalence : 0.6506
##       Balanced Accuracy : 0.9878
##
##        'Positive' Class : 2
##
```

```r
confusionMatrix(cm_test)
```

```
## Confusion Matrix and Statistics
##
##    test_pred
##       2   4
##   2 133   0
##   4   1  71
##
##                Accuracy : 0.9951
##                  95% CI : (0.9731, 0.9999)
##     No Information Rate : 0.6537
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9893
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9925
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.9861
##              Prevalence : 0.6537
##          Detection Rate : 0.6488
```

```
##    Detection Prevalence : 0.6488
##       Balanced Accuracy : 0.9963
##
##        'Positive' Class : 2
##
```

Some information on the model

```
summary(classifier)
```

```
##
## Call:
## svm(formula = Class ~ ., data = data, kernel = "linear")
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  1
##
## Number of Support Vectors:  81
##
##  ( 36 45 )
##
##
## Number of Classes:  2
##
## Levels:
##  2 4
```