CS513 HW4: NB

Aidan Fischer - 10447681

2023-10-16

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
```

```
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)
```

Naive Bayes

```
classifier <- naiveBayes(Class ~ (.-Sample), data=train)
classifier
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##         2         4
## 0.6506276 0.3493724
##
```

```
## Conditional probabilities:
##    F1
## Y             1           2           3           4           5           6
##   2 0.299035370 0.109324759 0.215434084 0.154340836 0.186495177 0.028938907
##   4 0.005988024 0.023952096 0.053892216 0.047904192 0.197604790 0.065868263
##    F1
## Y             7           8           9          10
##   2 0.003215434 0.003215434 0.000000000 0.000000000
##   4 0.083832335 0.155688623 0.059880240 0.305389222
##
##    F2
## Y             1           2           3           4           5           6
##   2 0.839228296 0.077170418 0.057877814 0.016077170 0.000000000 0.000000000
##   4 0.023952096 0.023952096 0.107784431 0.113772455 0.131736527 0.095808383
##    F2
## Y             7           8           9          10
##   2 0.003215434 0.003215434 0.003215434 0.000000000
##   4 0.089820359 0.125748503 0.029940120 0.257485030
##
##    F3
## Y             1           2           3           4           5           6
##   2 0.768488746 0.122186495 0.070739550 0.022508039 0.003215434 0.003215434
##   4 0.011976048 0.029940120 0.113772455 0.113772455 0.137724551 0.107784431
##    F3
## Y             7           8           9          10
##   2 0.006430868 0.003215434 0.000000000 0.000000000
##   4 0.113772455 0.125748503 0.041916168 0.203592814
##
##    F4
## Y             1           2           3           4           5           6
##   2 0.797427653 0.090032154 0.077170418 0.012861736 0.009646302 0.009646302
##   4 0.107784431 0.089820359 0.137724551 0.131736527 0.089820359 0.083832335
##    F4
## Y             7           8           9          10
##   2 0.000000000 0.000000000 0.000000000 0.003215434
##   4 0.059880240 0.101796407 0.017964072 0.179640719
##
##    F5
## Y             1           2           3           4           5           6
##   2 0.086816720 0.810289389 0.070739550 0.009646302 0.012861736 0.003215434
##   4 0.005988024 0.101796407 0.155688623 0.221556886 0.101796407 0.167664671
##    F5
## Y             7           8           9          10
##   2 0.006430868 0.000000000 0.000000000 0.000000000
##   4 0.035928144 0.077844311 0.011976048 0.119760479
##
##    F6
## Y             1           2           3           4           5           6
##   2 0.861736334 0.045016077 0.028938907 0.019292605 0.028938907 0.000000000
##   4 0.071856287 0.029940120 0.047904192 0.053892216 0.107784431 0.011976048
##    F6
## Y             7           8           9          10
##   2 0.000000000 0.006430868 0.000000000 0.009646302
##   4 0.029940120 0.071856287 0.041916168 0.532934132
```

```
##
##    F7
## Y             1           2           3           4           5           6
##   2 0.347266881 0.334405145 0.286173633 0.016077170 0.006430868 0.000000000
##   4 0.011976048 0.029940120 0.131736527 0.113772455 0.131736527 0.041916168
##    F7
## Y             7           8           9          10
##   2 0.009646302 0.000000000 0.000000000 0.000000000
##   4 0.293413174 0.113772455 0.059880240 0.071856287
##
##    F8
## Y             1           2           3           4           5           6
##   2 0.890675241 0.061093248 0.022508039 0.003215434 0.006430868 0.006430868
##   4 0.173652695 0.023952096 0.107784431 0.071856287 0.071856287 0.077844311
##    F8
## Y             7           8           9          10
##   2 0.006430868 0.003215434 0.000000000 0.000000000
##   4 0.041916168 0.095808383 0.071856287 0.263473054
##
##    F9
## Y             1           2           3           4           5           6
##   2 0.971061093 0.012861736 0.006430868 0.000000000 0.003215434 0.000000000
##   4 0.550898204 0.119760479 0.137724551 0.041916168 0.017964072 0.011976048
##    F9
## Y             7           8           9          10
##   2 0.003215434 0.003215434 0.000000000 0.000000000
##   4 0.029940120 0.029940120 0.000000000 0.059880240
```

Evaluation

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

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

```
confusionMatrix(cm_train)
```

```
## Confusion Matrix and Statistics
##
##    train_pred
##       2   4
##   2 302   9
##   4   3 164
##
##                Accuracy : 0.9749
##                  95% CI : (0.9566, 0.987)
##     No Information Rate : 0.6381
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9452
##
##  Mcnemar's Test P-Value : 0.1489
##
##             Sensitivity : 0.9902
```

```
##             Specificity : 0.9480
##          Pos Pred Value : 0.9711
##          Neg Pred Value : 0.9820
##              Prevalence : 0.6381
##          Detection Rate : 0.6318
##    Detection Prevalence : 0.6506
##       Balanced Accuracy : 0.9691
##
##        'Positive' Class : 2
##
```

```
confusionMatrix(cm_test)
```

```
## Confusion Matrix and Statistics
##
##    test_pred
##       2   4
##   2 129   4
##   4   0  72
##
##                Accuracy : 0.9805
##                  95% CI : (0.9508, 0.9947)
##     No Information Rate : 0.6293
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9577
##
##  Mcnemar's Test P-Value : 0.1336
##
##             Sensitivity : 1.0000
##             Specificity : 0.9474
##          Pos Pred Value : 0.9699
##          Neg Pred Value : 1.0000
##              Prevalence : 0.6293
##          Detection Rate : 0.6293
##    Detection Prevalence : 0.6488
##       Balanced Accuracy : 0.9737
##
##        'Positive' Class : 2
##
```