HW4.R

Fumonchu

2021-10-21

Company : Stevens Project : HW4 Purpose : HW4 First Name : David Last Name : Fu CWID : 10471854 Date : October 20, 2021

Homework 4

```
rm(list=ls())
library(e1071)
library(class)
set.seed(513)
```

```
cancerData=read.csv("C:/Users/Fumonchu/Documents/GitHub/School/CS513/HW3/breast-cancer-wisconsi
n.csv", header=TRUE)
cancerData$F6 <- as.numeric(cancerData$F6)
```

```
## Warning: NAs introduced by coercion
```

```
cancerData <- cancerData[rowSums(is.na(cancerData)) == 0,]
cancerData <- na.omit(cancerData)
summary(cancerData)
```

```
##      Sample               F1               F2               F3
##  Min.   :   63375   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.:  877617   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000
##  Median : 1171795   Median : 4.000   Median : 1.000   Median : 1.000
##  Mean   : 1076720   Mean   : 4.442   Mean   : 3.151   Mean   : 3.215
##  3rd Qu.: 1238705   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000
##  Max.   :13454352   Max.   :10.000   Max.   :10.000   Max.   :10.000
##        F4              F5               F6               F7
##  Min.   : 1.00   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.: 1.00   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 2.000
##  Median : 1.00   Median : 2.000   Median : 1.000   Median : 3.000
##  Mean   : 2.83   Mean   : 3.234   Mean   : 3.545   Mean   : 3.445
##  3rd Qu.: 4.00   3rd Qu.: 4.000   3rd Qu.: 6.000   3rd Qu.: 5.000
##  Max.   :10.00   Max.   :10.000   Max.   :10.000   Max.   :10.000
##        F8              F9             Class
##  Min.   : 1.00   Min.   : 1.000   Min.   :2.0
##  1st Qu.: 1.00   1st Qu.: 1.000   1st Qu.:2.0
##  Median : 1.00   Median : 1.000   Median :2.0
##  Mean   : 2.87   Mean   : 1.603   Mean   :2.7
##  3rd Qu.: 4.00   3rd Qu.: 1.000   3rd Qu.:4.0
##  Max.   :10.00   Max.   :10.000   Max.   :4.0
```

```
idx<-sort(sample(nrow(cancerData),as.integer(.70*nrow(cancerData))))

training<-cancerData[idx,]

test<-cancerData[-idx,]

nBayes_all <- naiveBayes(Class ~., data = training)

category_all<-predict(nBayes_all,test  )


table(NBayes_all=category_all,Class=test$Class)
```

```
##           Class
## NBayes_all   2   4
##          2 133   2
##          4   7  63
```

```
NB_wrong<-sum(category_all!=test$Class)

NB_error_rate<-NB_wrong/length(category_all)

rm(list=ls())
```