ML Session 4-NaiveBayes

Subiksha

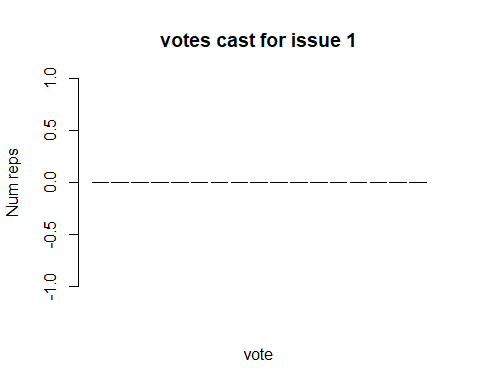
01/10/2020

library(mlbench)

## Warning: package 'mlbench' was built under R version 3.6.3

data("HouseVotes84")

plot(as.factor(HouseVotes84))  
title(main = 'votes cast for issue 1',xlab = "vote",ylab = "Num reps")



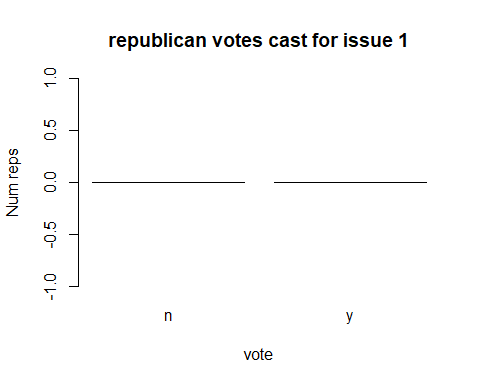
HouseVotes84$Class

## [1] republican republican democrat democrat democrat democrat   
## [7] democrat republican republican democrat republican republican  
## [13] democrat democrat republican republican democrat democrat   
## [19] republican democrat democrat democrat democrat democrat   
## [25] democrat democrat democrat democrat republican democrat   
## [31] republican democrat democrat republican democrat republican  
## [37] republican republican republican democrat democrat democrat   
## [43] democrat democrat democrat democrat democrat democrat   
## [49] democrat republican democrat republican democrat republican  
## [55] democrat republican republican republican republican republican  
## [61] democrat republican democrat democrat democrat republican  
## [67] republican republican democrat democrat democrat republican  
## [73] democrat republican democrat democrat democrat democrat   
## [79] democrat republican democrat democrat republican republican  
## [85] republican democrat republican republican democrat republican  
## [91] democrat democrat democrat democrat democrat democrat   
## [97] democrat democrat democrat republican democrat democrat   
## [103] democrat democrat democrat democrat republican republican  
## [109] democrat democrat democrat republican democrat republican  
## [115] democrat democrat democrat republican democrat republican  
## [121] republican republican republican republican democrat republican  
## [127] republican democrat democrat democrat democrat democrat   
## [133] republican republican republican republican republican democrat   
## [139] democrat democrat republican republican republican democrat   
## [145] democrat democrat republican democrat republican democrat   
## [151] republican democrat democrat democrat republican republican  
## [157] republican democrat republican democrat democrat democrat   
## [163] democrat republican democrat democrat republican republican  
## [169] democrat democrat democrat republican democrat democrat   
## [175] democrat democrat republican democrat democrat democrat   
## [181] democrat democrat democrat democrat democrat democrat   
## [187] democrat democrat republican democrat republican republican  
## [193] democrat democrat democrat republican democrat republican  
## [199] democrat democrat democrat democrat democrat democrat   
## [205] republican democrat republican republican democrat democrat   
## [211] democrat republican democrat democrat republican democrat   
## [217] democrat republican democrat democrat democrat democrat   
## [223] democrat republican republican republican democrat democrat   
## [229] republican republican republican republican democrat republican  
## [235] democrat republican democrat democrat democrat republican  
## [241] republican democrat republican democrat democrat democrat   
## [247] democrat republican republican democrat republican republican  
## [253] democrat republican democrat democrat republican republican  
## [259] democrat democrat democrat democrat democrat democrat   
## [265] democrat democrat republican republican democrat democrat   
## [271] democrat democrat democrat republican republican democrat   
## [277] republican republican republican republican democrat republican  
## [283] republican republican democrat democrat democrat democrat   
## [289] democrat democrat democrat democrat democrat democrat   
## [295] democrat republican republican democrat democrat democrat   
## [301] republican democrat republican republican republican republican  
## [307] republican democrat republican democrat republican democrat   
## [313] democrat republican republican republican democrat democrat   
## [319] democrat democrat democrat democrat democrat democrat   
## [325] republican democrat democrat republican democrat democrat   
## [331] republican democrat democrat democrat democrat republican  
## [337] democrat democrat democrat republican republican democrat   
## [343] democrat republican democrat republican republican republican  
## [349] democrat republican democrat republican democrat republican  
## [355] democrat republican republican republican democrat republican  
## [361] democrat democrat democrat republican republican democrat   
## [367] democrat democrat democrat republican democrat democrat   
## [373] democrat democrat republican democrat democrat republican  
## [379] republican republican democrat democrat democrat democrat   
## [385] democrat democrat democrat democrat democrat democrat   
## [391] democrat democrat republican republican democrat democrat   
## [397] democrat democrat democrat republican republican republican  
## [403] republican republican republican republican democrat democrat   
## [409] democrat republican republican democrat republican republican  
## [415] democrat democrat republican democrat democrat democrat   
## [421] republican democrat democrat democrat democrat democrat   
## [427] democrat republican democrat democrat republican democrat   
## [433] republican republican republican  
## Levels: democrat republican

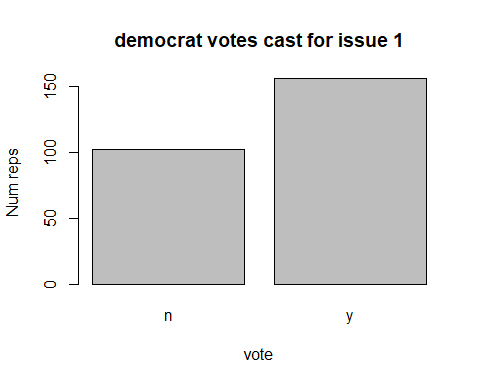
repub <- HouseVotes84$Class=="republion"  
democrat <- HouseVotes84$Class=="democrat"  
repub

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [217] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [229] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [241] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [253] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [265] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [277] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [289] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [301] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [313] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [325] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [337] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [349] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [361] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [373] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [385] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [397] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [409] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [421] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [433] FALSE FALSE FALSE

plot(as.factor(HouseVotes84[repub,2]))  
title(main = "republican votes cast for issue 1",xlab = "vote",ylab = "Num reps")



plot(as.factor(HouseVotes84[democrat,2]))  
title(main = "democrat votes cast for issue 1",xlab = "vote",ylab = "Num reps")



na\_by\_col\_class <- function(col,cls){return(sum(is.na(HouseVotes84[,col]) & HouseVotes84$Class==cls))}  
na\_by\_col\_class

## function(col,cls){return(sum(is.na(HouseVotes84[,col]) & HouseVotes84$Class==cls))}

p\_y\_col\_class <- function(col,cls){   
 sum\_y <- sum(HouseVotes84[,col]=="y" & HouseVotes84$Class==cls,na.rm = TRUE)   
 sum\_n <- sum(HouseVotes84[,col]=="n"& HouseVotes84$Class==cls,na.rm = TRUE)   
 return(sum\_y/(sum\_y+sum\_n))  
 }

p\_y\_col\_class(5,"democrat")

## [1] 0.05405405

p\_y\_col\_class(5,"republican")

## [1] 0.9878788

for(i in 2:ncol(HouseVotes84)){  
 if(sum(is.na(HouseVotes84[,i])>0)){  
 c1 <- which(is.na(HouseVotes84[,i])& HouseVotes84$Class=="democrat",arr.ind = TRUE)  
 c2 <- which(is.na(HouseVotes84[,i])& HouseVotes84$Class=="republican",arr.ind = TRUE)  
 HouseVotes84[c1,i] <-   
 ifelse(runif(na\_by\_col\_class(i,"democrat"))<p\_y\_col\_class(i,"democrat"),"y","n")  
 HouseVotes84[c2,i] <-   
 ifelse(runif(na\_by\_col\_class(i,"republican"))<p\_y\_col\_class(i,"republican"),"y","n")}  
}

HouseVotes84[,"train"] <- ifelse(runif(nrow(HouseVotes84))<0.80,1,0)  
traincolNum <- grep("train",names(HouseVotes84))  
trainhousevotes84 <- HouseVotes84[HouseVotes84$train==1,traincolNum]  
trainhousevotes84 <- HouseVotes84[HouseVotes84$train==0,traincolNum]

library(e1071)

## Warning: package 'e1071' was built under R version 3.6.3