Assign2-1\_Naive Bayes

Jonathan De Los Santos

2/6/2021

# Prompt

Calculate the conditional probability of the “Delay” given “Carrier = DL,” “Day of Week=Saturday (7),” “Destination = LGA,” and “Origin = DCA.” (Show your work by writing both equations and computation in R.)

# Data Prep

Read in csv, convert attributes to factors

delay <- read.csv("Data Sets/3.1-Delay.csv")  
str(delay)

## 'data.frame': 428 obs. of 5 variables:  
## $ Day.of.Week: int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Carrier : chr "DH" "US" "DH" "DH" ...  
## $ Origin : chr "IAD" "DCA" "IAD" "IAD" ...  
## $ Destination: chr "LGA" "LGA" "LGA" "LGA" ...  
## $ Status : int 0 0 0 0 0 0 0 0 1 0 ...

# Create column vector for factor conversion  
columns <- c(1:5)  
  
# Apply factor to all columns  
delay[,columns] <- lapply(delay[,columns], factor)

# Naive Bayes Model

Using the naiveBayes() function from the e1071 package

library(e1071)  
nb.model <- naiveBayes(Status~ ., data = delay)  
nb.model

##   
## Naive Bayes Classifier for Discrete Predictors  
##   
## Call:  
## naiveBayes.default(x = X, y = Y, laplace = laplace)  
##   
## A-priori probabilities:  
## Y  
## 0 1   
## 0.92523364 0.07476636   
##   
## Conditional probabilities:  
## Day.of.Week  
## Y 1 2 3 4 5 6  
## 0 0.17676768 0.12121212 0.14393939 0.14393939 0.18686869 0.06060606  
## 1 0.43750000 0.46875000 0.00000000 0.00000000 0.03125000 0.00000000  
## Day.of.Week  
## Y 7  
## 0 0.16666667  
## 1 0.06250000  
##   
## Carrier  
## Y CO DH DL MQ OH RU  
## 0 0.060606061 0.318181818 0.111111111 0.176767677 0.010101010 0.229797980  
## 1 0.062500000 0.343750000 0.093750000 0.312500000 0.000000000 0.093750000  
## Carrier  
## Y UA US  
## 0 0.007575758 0.085858586  
## 1 0.062500000 0.031250000  
##   
## Origin  
## Y BWI DCA IAD  
## 0 0.09090909 0.51515152 0.39393939  
## 1 0.03125000 0.53125000 0.43750000  
##   
## Destination  
## Y EWR JFK LGA  
## 0 0.3863636 0.1919192 0.4217172  
## 1 0.2500000 0.2500000 0.5000000

# Conditional Probability Calculations

## Save Probability Values

Calculate probability of “Carrier = DL,” “Day of Week=Saturday (7),” “Destination = LGA,” and “Origin = DCA” given “Status” of both 0 and 1.

DL.1 = .094   
DL.0 = .111  
  
Sat.1 = 0.0625  
Sat.0 = 0.1667  
  
LGA.1 = 0.5  
LGA.0 = 0.422  
  
DCA.1 = 0.531  
DCA.0 = 0.515  
  
Status.1 = 0.075  
Status.0 = 0.925  
  
pDelay = DL.1\*Sat.1\*LGA.1\*DCA.1\*Status.1  
cat("P(DL,Sat,LGA,DCA | Status = 1) = ", pDelay,"\n")

## P(DL,Sat,LGA,DCA | Status = 1) = 0.0001169859

pNoDelay = DL.0\*Sat.0\*LGA.0\*DCA.0\*Status.0  
cat("P(DL,Sat,LGA,DCA | Status = 0) = ", pNoDelay)

## P(DL,Sat,LGA,DCA | Status = 0) = 0.003719803

## Calculate Conditional Probability

The conditional probability of a delay given the specified carrier, day of week, destination, and origin is:

cpDelay = pDelay/(pDelay+pNoDelay)  
cpDelay

## [1] 0.03049058