Q7

Niraj Raj Kharel (16)

2024-05-31

Question 7

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.3
```

```
data("Arrests")
crime <- Arrests[, -8]
str(crime)
```

```
## 'data.frame':    5226 obs. of  7 variables:
##  $ released: Factor w/ 2 levels "No","Yes": 2 1 2 1 2 2 2 2 2 2 ...
##  $ colour  : Factor w/ 2 levels "Black","White": 2 1 2 1 1 1 2 2 1 2 ...
##  $ year    : int  2002 1999 2000 2000 1999 1998 1999 1998 2000 2001 ...
##  $ age     : int  21 17 24 46 27 16 40 34 23 30 ...
##  $ sex     : Factor w/ 2 levels "Female","Male": 2 2 2 2 1 1 2 1 2 2 ...
##  $ employed: Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 1 2 2 2 ...
##  $ citizen : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
```

```
# Question a
# Split the dataset into training and testing dataset
set.seed(16)
index <- sample(2, nrow(crime), replace = T, prob = c(0.8, 0.2))
train.crime = crime[index == 1, ] # Training dataset
test.crime = crime[index == 2, ] # Testing dataset

# Question b
# Logistic Regression
model_lr <- glm(released ~ ., data = train.crime, family = binomial)

# Naive Bayes
library(e1071)
```

```
## Warning: package 'e1071' was built under R version 4.3.3
```

```
set.seed(16)
model_nb<-naiveBayes(released ~., data= train.crime)

# Question c
# For LR
predicted_lr <- predict(model_lr, test.crime)
```

```r
predict_lr<-as.numeric((ifelse(predicted_lr>0.5,1,0)))
conf_matrix <- table(test.crime$released, predict_lr)
print(conf_matrix)
```

```
##      predict_lr
##         0   1
##   No   18 158
##   Yes  23 842
```

```r
# Calculate sensitivity and specificity
TN <- conf_matrix[1, 1]
FP <- conf_matrix[1, 2]
FN <- conf_matrix[2, 1]
TP <- conf_matrix[2, 2]

(sensitivity <- TP / (TP + FN))
```

```
## [1] 0.9734104
```

```r
(specificity <- TN / (TN + FP))
```

```
## [1] 0.1022727
```

```r
(accuracy_lr<-sum(diag(conf_matrix)/sum(conf_matrix)))
```

```
## [1] 0.8261287
```

```r
#For Naive Bayes

predict_nb<-predict(model_nb,newdata=test.crime)
cm_nb<-table(test.crime$released,predict_nb)
cm_nb
```

```
##      predict_nb
##        No Yes
##   No   15 161
##   Yes  19 846
```

```r
(accuracy_nb<-sum(diag(cm_nb)/sum(cm_nb)))
```

```
## [1] 0.8270893
```

```r
# Question d
# Here both the classification model show similar result of accuracy.
# Therefore, any one of the method can be employed for
#classification.
```