Assignment

Rohit

2024-02-20

library(ISLR)  
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

# Load Carseats dataset  
data(Carseats)

# View the dataset and get summary statistics  
View(Carseats)  
summary(Carseats)

## Sales CompPrice Income Advertising   
## Min. : 0.000 Min. : 77 Min. : 21.00 Min. : 0.000   
## 1st Qu.: 5.390 1st Qu.:115 1st Qu.: 42.75 1st Qu.: 0.000   
## Median : 7.490 Median :125 Median : 69.00 Median : 5.000   
## Mean : 7.496 Mean :125 Mean : 68.66 Mean : 6.635   
## 3rd Qu.: 9.320 3rd Qu.:135 3rd Qu.: 91.00 3rd Qu.:12.000   
## Max. :16.270 Max. :175 Max. :120.00 Max. :29.000   
## Population Price ShelveLoc Age Education   
## Min. : 10.0 Min. : 24.0 Bad : 96 Min. :25.00 Min. :10.0   
## 1st Qu.:139.0 1st Qu.:100.0 Good : 85 1st Qu.:39.75 1st Qu.:12.0   
## Median :272.0 Median :117.0 Medium:219 Median :54.50 Median :14.0   
## Mean :264.8 Mean :115.8 Mean :53.32 Mean :13.9   
## 3rd Qu.:398.5 3rd Qu.:131.0 3rd Qu.:66.00 3rd Qu.:16.0   
## Max. :509.0 Max. :191.0 Max. :80.00 Max. :18.0   
## Urban US   
## No :118 No :142   
## Yes:282 Yes:258   
##   
##   
##   
##

# Establish a seed for reproducibility  
set.seed(123)

# Finding out the total number of rows in the dataset.  
total\_rows <- nrow(Carseats)

# Determine the quantity of rows in the training, validation, and test sets.  
training\_set <- round(0.6 \* total\_rows)  
validation\_set <- round(0.2 \* total\_rows)  
Test\_num <- total\_rows - training\_set - validation\_set

# Use the same seed to ensure reproducibility and shuffle the dataset at random.  
shuffled\_indices <- sample(seq\_len(total\_rows))

# Create a training set using 60% of the data.  
Training <- Carseats[shuffled\_indices[1:training\_set], ]

# Produce a validation set using 20% of the data.  
Validation <- Carseats[shuffled\_indices[(training\_set + 1):(training\_set + validation\_set)], ]

# Generate a test set (20% of the data).  
Test <- Carseats[shuffled\_indices[(training\_set + validation\_set + 1):total\_rows], ]

# Verify if there are duplicates.  
if (any(duplicated(Training) | duplicated(Validation) | duplicated(Test))) {  
 stop("Duplicate records found!")  
}

# Check the dimensions of the datasets  
dim(Training)

## [1] 240 11

dim(Validation)

## [1] 80 11

dim(Test)

## [1] 80 11