Advanced DM Assignment two final

Khutso Ledwaba

2023-03-21

```
library('ISLR')
```

```
## Warning: package 'ISLR' was built under R version 4.2.2
```

```
library('dplyr')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library('glmnet')
```

```
## Warning: package 'glmnet' was built under R version 4.2.2
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.2.2
```

```
## Loaded glmnet 4.1-6
```

```
library('caret')
```

```
## Loading required package: ggplot2
```

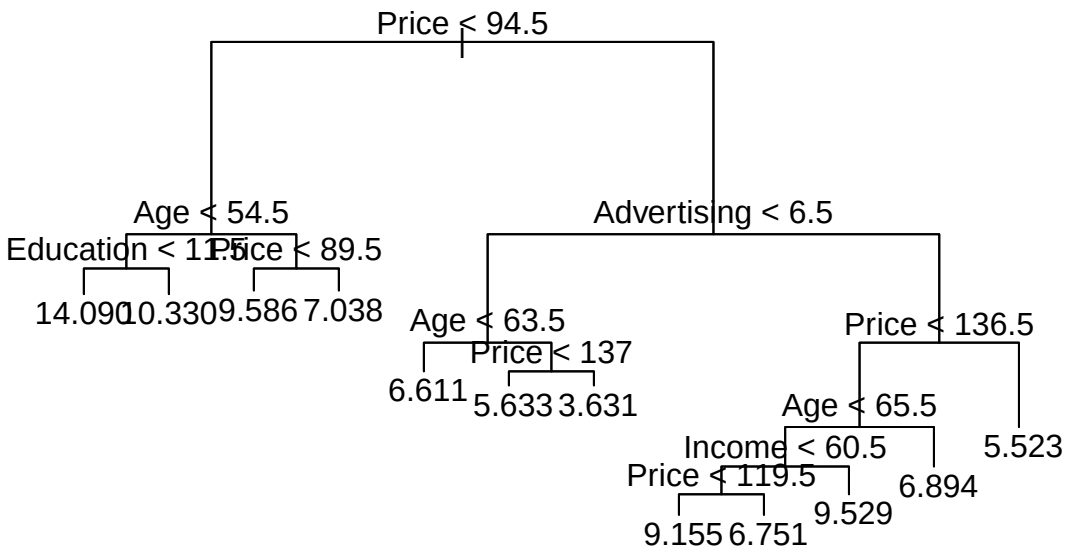
```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Loading required package: lattice
```

```
library('tree')
```

```
## Warning: package 'tree' was built under R version 4.2.3
```

```
library('class')


Carseats <- Carseats

Carseats_Filtered <- Carseats %>% select("Sales", "Price", "Advertising","Population","Age","Income","E

##Question 1

# Build the decision tree regression model
tree.sales <- tree(Sales ~ ., data = Carseats_Filtered)

# Plot the decision tree
plot(tree.sales)
text(tree.sales, pretty = 0)
```




```
##Question 2

# Create a data frame with the input record
newdata <- data.frame(Sales = 9, Price = 6.54, Population = 124,Advertising = 0,Age = 76,Income = 110,E
```

```r
# Use the predict() function to estimate the Sales
predicted.sales <- predict(tree.sales, newdata = newdata)

# Print the estimated Sales
cat("The estimated Sales for the given input record is ", round(predicted.sales, 3), ".")
```

```
## The estimated Sales for the given input record is  9.586 .
```

```r
##Question 3
#We will use the train() function from the caret package to train a random forest model for the carseat


rf.sales <- train(Sales ~ .,data = Carseats_Filtered, method = "rf")

#Finding the Optimal Value of mtry
rf.sales
```

```
## Random Forest
##
## 400 samples
##   6 predictor
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 400, 400, 400, 400, 400, 400, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##   2     2.408475  0.2774283  1.918575
##   4     2.429651  0.2710479  1.930468
##   6     2.458115  0.2605994  1.951537
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 2.
```

```r
##Question 4

#Customizing the Search Grid

# Create the search grid
grid <- expand.grid(mtry = c(2, 3, 5))

# Specify the cross-validation settings
ctrl <- trainControl(
  method = "repeatedcv",
  repeats = 3,
  number = 5,
  verboseIter = TRUE
)

# Train the model with the customized search grid
rf.sales <- train(
```

```
  Sales ~ .,
  data = Carseats_Filtered,
  method = "rf",
  trControl = ctrl,
  tuneGrid = grid
)
```

```
## + Fold1.Rep1: mtry=2
## - Fold1.Rep1: mtry=2
## + Fold1.Rep1: mtry=3
## - Fold1.Rep1: mtry=3
## + Fold1.Rep1: mtry=5
## - Fold1.Rep1: mtry=5
## + Fold2.Rep1: mtry=2
## - Fold2.Rep1: mtry=2
## + Fold2.Rep1: mtry=3
## - Fold2.Rep1: mtry=3
## + Fold2.Rep1: mtry=5
## - Fold2.Rep1: mtry=5
## + Fold3.Rep1: mtry=2
## - Fold3.Rep1: mtry=2
## + Fold3.Rep1: mtry=3
## - Fold3.Rep1: mtry=3
## + Fold3.Rep1: mtry=5
## - Fold3.Rep1: mtry=5
## + Fold4.Rep1: mtry=2
## - Fold4.Rep1: mtry=2
## + Fold4.Rep1: mtry=3
## - Fold4.Rep1: mtry=3
## + Fold4.Rep1: mtry=5
## - Fold4.Rep1: mtry=5
## + Fold5.Rep1: mtry=2
## - Fold5.Rep1: mtry=2
## + Fold5.Rep1: mtry=3
## - Fold5.Rep1: mtry=3
## + Fold5.Rep1: mtry=5
## - Fold5.Rep1: mtry=5
## + Fold1.Rep2: mtry=2
## - Fold1.Rep2: mtry=2
## + Fold1.Rep2: mtry=3
## - Fold1.Rep2: mtry=3
## + Fold1.Rep2: mtry=5
## - Fold1.Rep2: mtry=5
## + Fold2.Rep2: mtry=2
## - Fold2.Rep2: mtry=2
## + Fold2.Rep2: mtry=3
## - Fold2.Rep2: mtry=3
## + Fold2.Rep2: mtry=5
## - Fold2.Rep2: mtry=5
## + Fold3.Rep2: mtry=2
## - Fold3.Rep2: mtry=2
## + Fold3.Rep2: mtry=3
## - Fold3.Rep2: mtry=3
```

```
## + Fold3.Rep2: mtry=5
## - Fold3.Rep2: mtry=5
## + Fold4.Rep2: mtry=2
## - Fold4.Rep2: mtry=2
## + Fold4.Rep2: mtry=3
## - Fold4.Rep2: mtry=3
## + Fold4.Rep2: mtry=5
## - Fold4.Rep2: mtry=5
## + Fold5.Rep2: mtry=2
## - Fold5.Rep2: mtry=2
## + Fold5.Rep2: mtry=3
## - Fold5.Rep2: mtry=3
## + Fold5.Rep2: mtry=5
## - Fold5.Rep2: mtry=5
## + Fold1.Rep3: mtry=2
## - Fold1.Rep3: mtry=2
## + Fold1.Rep3: mtry=3
## - Fold1.Rep3: mtry=3
## + Fold1.Rep3: mtry=5
## - Fold1.Rep3: mtry=5
## + Fold2.Rep3: mtry=2
## - Fold2.Rep3: mtry=2
## + Fold2.Rep3: mtry=3
## - Fold2.Rep3: mtry=3
## + Fold2.Rep3: mtry=5
## - Fold2.Rep3: mtry=5
## + Fold3.Rep3: mtry=2
## - Fold3.Rep3: mtry=2
## + Fold3.Rep3: mtry=3
## - Fold3.Rep3: mtry=3
## + Fold3.Rep3: mtry=5
## - Fold3.Rep3: mtry=5
## + Fold4.Rep3: mtry=2
## - Fold4.Rep3: mtry=2
## + Fold4.Rep3: mtry=3
## - Fold4.Rep3: mtry=3
## + Fold4.Rep3: mtry=5
## - Fold4.Rep3: mtry=5
## + Fold5.Rep3: mtry=2
## - Fold5.Rep3: mtry=2
## + Fold5.Rep3: mtry=3
## - Fold5.Rep3: mtry=3
## + Fold5.Rep3: mtry=5
## - Fold5.Rep3: mtry=5
## Aggregating results
## Selecting tuning parameters
## Fitting mtry = 2 on full training set
```

```
# Print the results
rf.sales
```

```
## Random Forest
##
## 400 samples
```

```
##   6 predictor
##
## No pre-processing
## Resampling: Cross-Validated (5 fold, repeated 3 times)
## Summary of sample sizes: 321, 320, 320, 320, 319, 320, ...
## Resampling results across tuning parameters:
##
##   mtry  RMSE      Rsquared   MAE
##   2     2.382558  0.2951713  1.907019
##   3     2.384506  0.2938716  1.903278
##   5     2.407822  0.2861416  1.920766
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was mtry = 2.
```