STAT 420: Homework 03

Chaitanya Sood - csood2

Due: Tuesday, September 29 by 11:30 PM CT

Contents

```
#install.packages("mltools")
library(mltools)
library(data.table)
```

Loading dataset in R from the CSV file, and converting

```
car_data = read.csv("Cars_data.csv")
car_data = subset(car_data, select = -c(Vehicle.Style, Market.Category))

#unique(car_data$Transmission.Type)
#colnames(car_data)
#unique(car_data$Engine.Fuel.Type)
#unique(car_data$Driven_Wheels)
unique(car_data$Vehicle.Size)
```

```
## [1] "Compact" "Midsize" "Large"
```

```
#head(car_data)
```

Removing the dew non automatic/manual transmission types, and storing this new data in car_data_transd dataframe

```
car_data_transd<-car_data[!(car_data$Transmission.Type=="AUTOMATED_MANUAL" | car_data$Transmission.Type=

unique(car_data_transd$Transmission.Type)
```

```
## [1] "MANUAL"    "AUTOMATIC"
```

Removing certain fuel types, keeping only gasoline and diesel. Storing the result in car_data_fuel dataframe

```
car_data_fuel<-car_data_transd[!(grepl("flex", car_data_transd$Engine.Fuel.Type, fixed = TRUE)
|car_data_transd$Engine.Fuel.Type=="electric" | car_data_transd$Engine.Fuel.Type=="" | car_data_transd$E

unique(car_data_fuel$Engine.Fuel.Type)
```

```
## [1] "premium unleaded (required)"    "regular unleaded"
## [3] "premium unleaded (recommended)" "diesel"
```

Assigning the different types of gasoline to a single "gasoline value". Now, the only two values
for fuel type will be "gasoline" and "diesel" as visible below

```
car_data_fuel$Engine.Fuel.Type[car_data_fuel$Engine.Fuel.Type == "premium unleaded (required)" ] <- "ga
car_data_fuel$Engine.Fuel.Type[car_data_fuel$Engine.Fuel.Type == "regular unleaded" ] <- "gasoline"
car_data_fuel$Engine.Fuel.Type[car_data_fuel$Engine.Fuel.Type == "premium unleaded (recommended)" ] <-

unique(car_data_fuel$Engine.Fuel.Type)
```

```
## [1] "gasoline" "diesel"
```

One hot encoding - converting categorical variables to values which R will understand

```
car_data_factored = car_data_fuel
car_data_factored$Vehicle.Size <- factor(car_data_factored$Vehicle.Size)
car_data_factored$Transmission.Type <- factor(car_data_factored$Transmission.Type)
car_data_factored$Engine.Fuel.Type <- factor(car_data_factored$Engine.Fuel.Type)
car_data_factored$Driven_Wheels <- factor(car_data_factored$Driven_Wheels)

levels(car_data_factored$Vehicle.Size)
```

```
## [1] "Compact" "Large"   "Midsize"
```

```
levels(car_data_factored$Transmission.Type)
```

```
## [1] "AUTOMATIC" "MANUAL"
```

```
levels(car_data_factored$Engine.Fuel.Type)
```

```
## [1] "diesel"   "gasoline"
```

```
levels(car_data_factored$Driven_Wheels)
```

```
## [1] "all wheel drive"   "four wheel drive"  "front wheel drive"
## [4] "rear wheel drive"
```

```
temp4 = one_hot(as.data.table(car_data_factored))
```