

1. کتابخانه‌های مورد نیاز برای پردازش داده‌ها بارگیری شدند، که شامل کتابخانه‌های `ggplot2`، `corrplot`، `tidyverse`، `fastDummies` و `caret` بودند.

2. `read.csv` ذخیره شدند. برای این کار، از تابع `data` مربوطه خوانده شده و در متغیر `csv` سپس داده‌های مورد نظر از فایل `read.csv` استفاده شده است `base` کتابخانه.

3. `ggplot` و `geom_point`، نمودارهایی برای بررسی همبستگی بین متغیرها ایجاد شده‌اند. برای این کار از تابع `ggplot2` با استفاده از کتابخانه `ggplot` استفاده شده است.

4. نمودار ماتریس همبستگی برای متغیرهای عددی ایجاد شده است. برای این کار از تابع `corrplot` با استفاده از کتابخانه `corrplot` استفاده شده است.

5. برخی از متغیرها به شکل عددی تبدیل شده‌اند تا در آنالیز همبستگی استفاده شوند. برای تبدیل متغیرهای دسته‌ای به عددی از تابع `as.numeric` استفاده شده است.

6. مقادیر خالی موجود در داده‌ها پر شده و متغیرهای دسته‌ای تبدیل به متغیرهای عددی شده‌اند. `tidyverse` با استفاده از کتابخانه `tidyverse` استفاده شده است `replace` و `na.omit` برای این کار از توابع `mutate`، `as.numeric` و `replace` استفاده شده است.

7. `dummy_cols`، متغیرهای دسته‌ای به متغیرهای دمی تبدیل شده‌اند. برای این کار از تابع `fastDummies` با استفاده از کتابخانه `fastDummies` استفاده شده است.

8. `createDataPartition`، داده‌ها به دو قسمت آموزشی و تست تقسیم شده‌اند. برای این کار از تابع `caret` با استفاده از کتابخانه `caret` استفاده شده است.

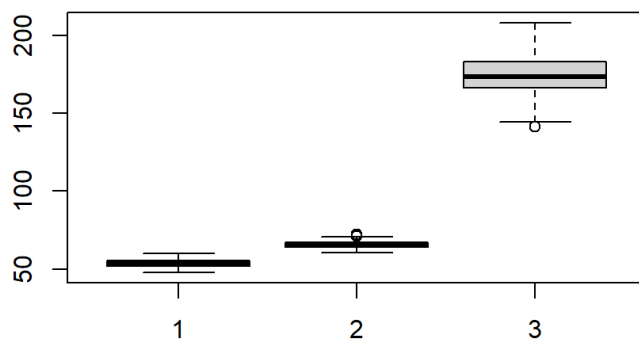
`data train and test` در بخش دوم به بررسی داده با رگرسیون چندگانه پرداختیم. و معیارهای مختلفی را بر روی بررسی کردیم.

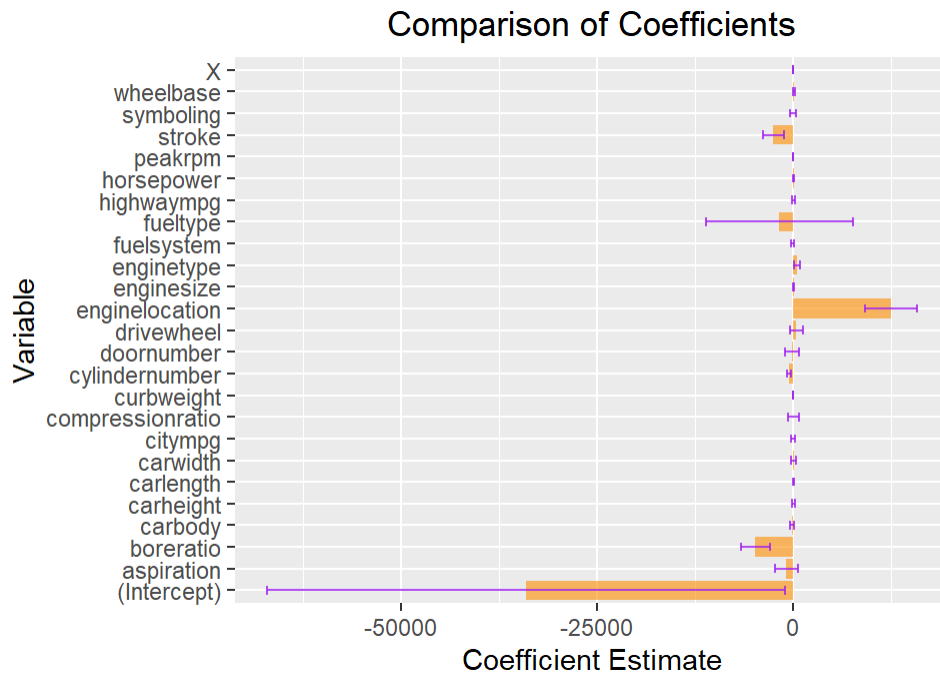
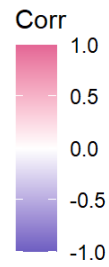
بخش آخر نیز یک مدل رگرسیون لجستیک روی داده‌های قیمت خودروها اجرا می‌کند. این داده شامل ویژگی‌های مختلف خودروها مانند نوع سوخت، تعداد سیلندر، جنس بدنه، موقعیت موتور و ... است. مدل رگرسیون لجستیک برای پیش‌بینی قیمت خودروها با استفاده از این ویژگی‌ها روی داده‌های آموزش آموزش داده شده است و سپس بر روی داده‌های آزمون ارزیابی شده است.

حذف شده، ردیف هایی که دارای مقادیر نامعتبر هستند حذف شده، car\_ID خوانده شده، ستون CSV در این کد، داده ها از یک فایل متغیرهای دسته‌ای به متغیرهای عددی تبدیل شده و متغیر قیمت به یک متغیر دودویی تبدیل شده است. همچنین مقادیر خالی با مقدار قبلی یا بعدی پر شده اند و متغیرهای رشته‌ای به متغیرهای فاکتور تبدیل شده اند

آموزش MASS از کتابخانه glm داده ها به دو مجموعه آموزش و آزمون تقسیم شده اند و مدل رگرسیون لجستیک با استفاده از تابع داده شده است. سپس معیارهای ارزیابی مدل برای داده های آموزش و آزمون محاسبه شده است

برای هر دو مجموعه داده (Specificity) و اختصاصیت (Sensitivity) ، حساسیت (Accuracy) نتایج ارزیابی شامل معیارهای دقت . چاپ شده است. این معیارها نشان می دهند که مدل چقدر دقیق در پیش بینی قیمت خودروها عمل کرده است





```

Call:
lm(formula = price ~ ., data = train_data)

Residuals:
    Min       1Q   Median       3Q      Max
-8971.9 -1411.4   -21.4  1225.8 10547.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.401e+04  3.302e+04  -1.030  0.305830
X            -1.600e+01  5.499e+00  -2.910  0.004558 **
symboling    -2.242e+01  3.564e+02  -0.063  0.949991
fueltype     -1.758e+03  9.350e+03  -0.188  0.851287
aspiration    -8.282e+02  1.451e+03  -0.571  0.569627
doornumber    -1.321e+02  8.575e+02  -0.154  0.877865
carbody       -9.619e+01  2.407e+02  -0.400  0.690365
drivewheel    4.458e+02  7.922e+02   0.563  0.575015
engine        1.254e+04  3.315e+03   3.784  0.000278 ***
wheelbase     1.642e+02  1.493e+02   1.099  0.274483
carlength     6.241e+01  7.746e+01   0.806  0.422527
carwidth      7.752e+01  3.362e+02   0.231  0.818158
carheight     4.274e+01  1.919e+02   0.223  0.824251
curbweight    2.095e+00  2.475e+00   0.847  0.399367
enginetype     5.159e+02  3.407e+02   1.514  0.133528
cylindernumber -4.870e+02  2.742e+02  -1.776  0.079104 .
enginesize     7.738e+01  2.659e+01   2.910  0.004550 **
fuelsystem    -2.688e+01  2.137e+02  -0.126  0.900185
boreratio     -4.805e+03  1.826e+03  -2.631  0.010027 *
stroke        -2.500e+03  1.315e+03  -1.902  0.060419 .
compressionratio 4.948e+01  6.716e+02   0.074  0.941437
horsepower     7.612e+01  2.728e+01   2.791  0.006418 **
peakrpm       1.147e+00  8.466e-01   1.354  0.179051
citympg       -1.226e+01  2.347e+02  -0.052  0.958462
highwaympg     2.505e+01  1.976e+02   0.127  0.899413
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2709 on 90 degrees of freedom
Multiple R-squared:  0.9014,    Adjusted R-squared:  0.8751
F-statistic: 34.27 on 24 and 90 DF,  p-value: < 2.2e-16

```

```

> cat("Test data:\n")
Test data:
> cat("RSS: ", test_RSS, "\n")
RSS: 559835763
> cat("TSS: ", test_TSS, "\n")
TSS: 4244480529
> cat("MSE: ", test_MSE, "\n")
MSE: 11196715
> cat("R-squared: ", test_R_Squared, "\n")
R-squared: 0.8681026
> cat("Adjusted R-squared: ", test_adjusted_R_Squared, "\n\n")
Adjusted R-squared: 0.7414812

```

```

> cat("Train data:\n")
Train data:
> cat("RSS: ", train_RSS, "\n")
RSS: 660701744
> cat("TSS: ", train_TSS, "\n")
TSS: 6698428551
> cat("MSE: ", train_MSE, "\n")
MSE: 5745233
> cat("R-squared: ", train_R_Squared, "\n")
R-squared: 0.9013647
> cat("Adjusted R-squared: ", train_adjusted_R_Squared, "\n")
Adjusted R-squared: 0.8750619
>

```

```

glm(formula = price_cat ~ ., family = binomial, data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.736e-05 -2.100e-08 -2.100e-08  2.100e-08  4.741e-05

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.105e+01  4.866e+06  0.000  1.000
X            -2.056e-02  3.549e+02  0.000  1.000
symboling    -3.933e+00  4.311e+04  0.000  1.000
fueltype      3.150e+02  1.217e+06  0.000  1.000
aspiration    1.663e+00  2.526e+05  0.000  1.000
doornumber    2.103e+01  2.498e+05  0.000  1.000
carbody       4.908e+00  3.226e+04  0.000  1.000
drivewheel    -3.926e+01  1.191e+05  0.000  1.000
enginelocation -4.212e+02  4.874e+05 -0.001  0.999
wheelbase     3.515e+00  2.056e+04  0.000  1.000
carlength     -2.889e+00  1.839e+04  0.000  1.000
carwidth      -1.235e+01  3.502e+04  0.000  1.000
carheight     -1.112e+00  2.587e+04  0.000  1.000
curbweight    1.126e-01  3.535e+02  0.000  1.000
enginetype    -5.123e+00  6.081e+04  0.000  1.000
cylindernumber 6.931e+00  4.901e+04  0.000  1.000
enginesize    -7.388e-01  5.606e+03  0.000  1.000
fuelsystem    -5.906e-01  2.294e+04  0.000  1.000
bore ratio    4.136e+01  3.494e+05  0.000  1.000
stroke        3.646e+01  2.380e+05  0.000  1.000
compressionratio 2.408e+01  8.730e+04  0.000  1.000
horsepower    -3.465e-01  6.340e+03  0.000  1.000
peakrpm       -1.894e-02  1.126e+02  0.000  1.000
citympg       -3.191e+00  6.191e+04  0.000  1.000
highwaympg     4.667e+00  4.514e+04  0.000  1.000
price         2.260e-02  1.427e+01  0.002  0.999

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.4860e+02  on 114  degrees of freedom
Residual deviance: 8.7073e-09  on  89  degrees of freedom
AIC: 52

Number of Fisher Scoring iterations: 25

```

```

<J>
> # چاپ معیارهای ارزیابی برای داده های آزمون
> cat("Test dataset:\n")
Test dataset:
> cat("Accuracy: ", accuracy, "\n")
Accuracy: 0.88
> cat("Sensitivity: ", sensitivity, "\n")
Sensitivity: 0.8888889
> cat("Specificity: ", specificity, "\n")
Specificity: 0.875
> # چاپ معیارهای ارزیابی برای داده های آموزش
> cat("Train dataset:\n")
Train dataset:
> cat("Accuracy: ", accuracy_train, "\n")
Accuracy: 1
> cat("Sensitivity: ", sensitivity_train, "\n")
Sensitivity: 1
> cat("Specificity: ", specificity_train, "\n")
Specificity: 1

```