multiml

September 24, 2024

[3]:

```
import pandas as pd
import math
import matplotlib.pyplot as plt
```

[4]:

```
df=pd.read_csv("CarPrice_Assignment.csv")
df.head(5)
```

[4]:

```
   car_ID  symboling                   CarName fueltype aspiration doornumber  \
0       1          3        alfa-romero giulia      gas        std        two
1       2          3       alfa-romero stelvio      gas        std        two
2       3          1  alfa-romero Quadrifoglio      gas        std        two
3       4          2               audi 100 ls      gas        std       four
4       5          2                audi 100ls      gas        std       four

       carbody drivewheel enginelocation  wheelbase  …  enginesize  \
0  convertible        rwd          front       88.6  …         130
1  convertible        rwd          front       88.6  …         130
2    hatchback        rwd          front       94.5  …         152
3        sedan        fwd          front       99.8  …         109
4        sedan        4wd          front       99.4  …         136

   fuelsystem  boreratio  stroke compressionratio horsepower  peakrpm citympg  \
0        mpfi       3.47    2.68              9.0        111     5000      21
1        mpfi       3.47    2.68              9.0        111     5000      21
2        mpfi       2.68    3.47              9.0        154     5000      19
3        mpfi       3.19    3.40             10.0        102     5500      24
4        mpfi       3.19    3.40              8.0        115     5500      18

   highwaympg    price
0          27  13495.0
1          27  16500.0
2          26  16500.0
3          30  13950.0
4          22  17450.0

[5 rows x 26 columns]
```

[5]:
```
df.dtypes==object
```

[5]:
```
car_ID              False
symboling           False
CarName              True
fueltype             True
aspiration           True
doornumber           True
carbody              True
drivewheel           True
enginelocation       True
wheelbase           False
carlength           False
carwidth            False
carheight           False
curbweight          False
enginetype           True
cylindernumber       True
enginesize          False
fuelsystem           True
boreratio           False
stroke              False
compressionratio    False
horsepower          False
peakrpm             False
citympg             False
highwaympg          False
price               False
dtype: bool
```

[6]:
```
df["enginetype"].unique()
```

[6]:
```
array(['dohc', 'ohcv', 'ohc', 'l', 'rotor', 'ohcf', 'dohcv'], dtype=object)
```

[7]:
```
df["carbody"].unique()
```

[7]:
```
array(['convertible', 'hatchback', 'sedan', 'wagon', 'hardtop'],
      dtype=object)
```

[8]:
```
df=pd.get_dummies(df,columns=['fueltype'])
```

[9]:
```
df.drop("fueltype_gas",axis=1)
```

[9]:
```
     car_ID  symboling                    CarName aspiration doornumber  \
0         1          3         alfa-romero giulia        std        two
1         2          3        alfa-romero stelvio        std        two
2         3          1   alfa-romero Quadrifoglio        std        two
```

```
3          4          2             audi 100 ls        std       four
4          5          2              audi 100ls        std       four
..       …        …                       …        …        …
200      201         -1          volvo 145e (sw)       std       four
201      202         -1             volvo 144ea      turbo       four
202      203         -1             volvo 244dl        std       four
203      204         -1               volvo 246      turbo       four
204      205         -1             volvo 264gl      turbo       four

         carbody drivewheel enginelocation  wheelbase  carlength  …  \
0    convertible        rwd          front       88.6      168.8  …
1    convertible        rwd          front       88.6      168.8  …
2      hatchback        rwd          front       94.5      171.2  …
3          sedan        fwd          front       99.8      176.6  …
4          sedan        4wd          front       99.4      176.6  …
..           …          …             …          …          …  …
200        sedan        rwd          front      109.1      188.8  …
201        sedan        rwd          front      109.1      188.8  …
202        sedan        rwd          front      109.1      188.8  …
203        sedan        rwd          front      109.1      188.8  …
204        sedan        rwd          front      109.1      188.8  …

    fuelsystem  boreratio  stroke compressionratio horsepower  peakrpm  \
0         mpfi       3.47    2.68              9.0        111     5000
1         mpfi       3.47    2.68              9.0        111     5000
2         mpfi       2.68    3.47              9.0        154     5000
3         mpfi       3.19    3.40             10.0        102     5500
4         mpfi       3.19    3.40              8.0        115     5500
..         …          …     …                …          …        …
200       mpfi       3.78    3.15              9.5        114     5400
201       mpfi       3.78    3.15              8.7        160     5300
202       mpfi       3.58    2.87              8.8        134     5500
203        idi       3.01    3.40             23.0        106     4800
204       mpfi       3.78    3.15              9.5        114     5400

   citympg  highwaympg    price  fueltype_diesel
0       21          27  13495.0            False
1       21          27  16500.0            False
2       19          26  16500.0            False
3       24          30  13950.0            False
4       18          22  17450.0            False
..     …          …       …             …
200     23          28  16845.0            False
201     19          25  19045.0            False
202     18          23  21485.0            False
203     26          27  22470.0             True
204     19          25  22625.0            False
```

```
[205 rows x 26 columns]
```

[10]:
```
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
label_encoder = preprocessing.LabelEncoder()
```

[13]:
```
df["enginetype"]=label_encoder.fit_transform(df["enginetype"])
df["carbody"]=label_encoder.fit_transform(df["carbody"])
```

[14]:
```
X=df[["horsepower","fueltype_diesel","enginesize","enginetype","carbody"]]
Y=df[["price"]]
```

[15]:
```
X
```

[15]:
```
     horsepower  fueltype_diesel  enginesize  enginetype  carbody
0           111            False         130           0        0
1           111            False         130           0        0
2           154            False         152           5        2
3           102            False         109           3        3
4           115            False         136           3        3
..          …                …          …           …      …
200         114            False         141           3        3
201         160            False         141           3        3
202         134            False         173           5        3
203         106             True         145           3        3
204         114            False         141           3        3

[205 rows x 5 columns]
```

[16]:
```
X_train, X_test, Y_train, Y_test = train_test_split(  X, Y, test_size=0.3)
```

[17]:
```
model=LinearRegression()
model.fit(X_train,Y_train)
```

[17]:
```
LinearRegression()
```

[18]:
```
y_pred=model.predict(X_test)
```

[19]:
```
print('mean_squared_error : ', mean_squared_error(Y_test, y_pred))
print('mean_absolute_error : ', mean_absolute_error(Y_test, y_pred))
print('root_mean_squared_error : ', math.sqrt(mean_absolute_error(Y_test,␣
  ↪y_pred)))
```

```
mean_squared_error :  11878155.527813977
```

```
mean_absolute_error :  2550.549082589818
root_mean_squared_error :  50.502961126946
```