Homework-2

April 23, 2020

1 Exercise 1

Computations.

When we have $d = 1, n = 1$:

$$F(\beta) = (y - x\beta)^2 + \lambda\beta^2 \tag{1}$$

$$\nabla F(\beta) = -2x(y - x\beta) + 2\lambda\beta \tag{2}$$

For $d = 1, n > 1$ (by linearity):

$$F(\beta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\beta)^2 + \lambda\|\beta\|_2^2 \tag{3}$$

$$\nabla F(\beta) = \frac{1}{n}\sum_{i=1}^{n} -2x_i(y_i - x_i\beta) + 2\lambda\beta \tag{4}$$

Finally, for $d > 1, n > 1$, we generalize and simplify:

$$\nabla F(\beta) = -2X^T\mathbf{y} + 2X^TX\beta + 2\lambda\beta \tag{5}$$

Loading & standardizing "Hitters" data.

Defining *computegrad* & *graddescent* functions.

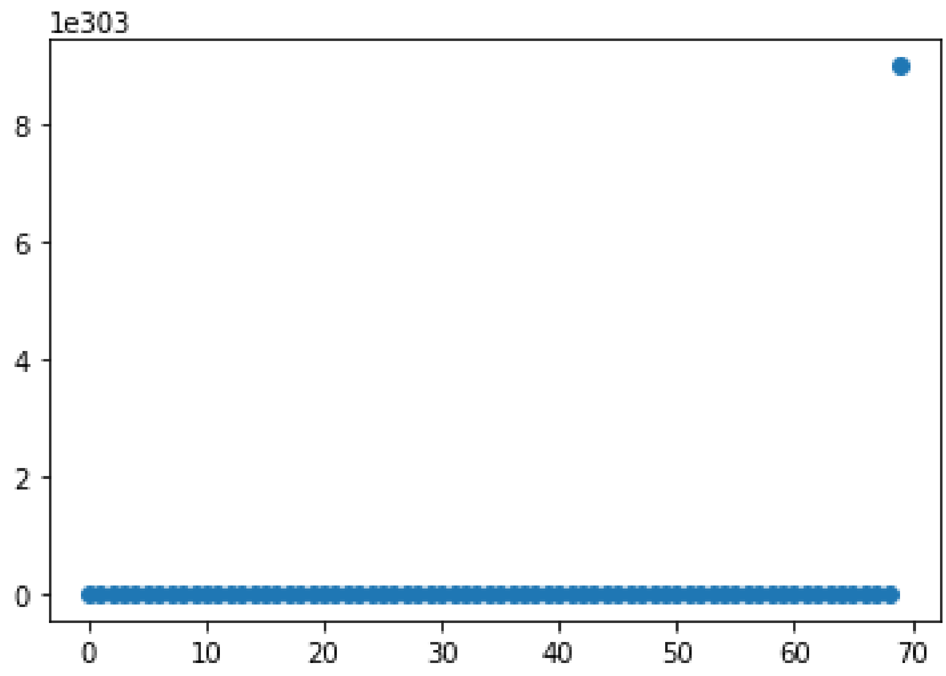
Using *graddescent* with $\eta = 0.05, t = 1000, \lambda = -5.00$, and plotting objective values.

We expect a negative λ value to overfit the data, since it grows the weights rather than penalizing the risk term. Accordingly, we should see our objective value (error) drastically tend towards $-\infty$.

Issue: I am instead seeing my objective function grow exponentially, eventually causing overflow errors and yielding NaN values. This implies an error in the objective function, or in the computation of B_t values, but both seem to match the derivation above.

```
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in double_scalars
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:5:
RuntimeWarning: invalid value encountered in double_scalars
  """
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:
RuntimeWarning: invalid value encountered in matmul

/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: invalid value encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: invalid value encountered in add
  after removing the cwd from sys.path.
```

```
<matplotlib.collections.PathCollection at 0x7f9513337cd0>
```



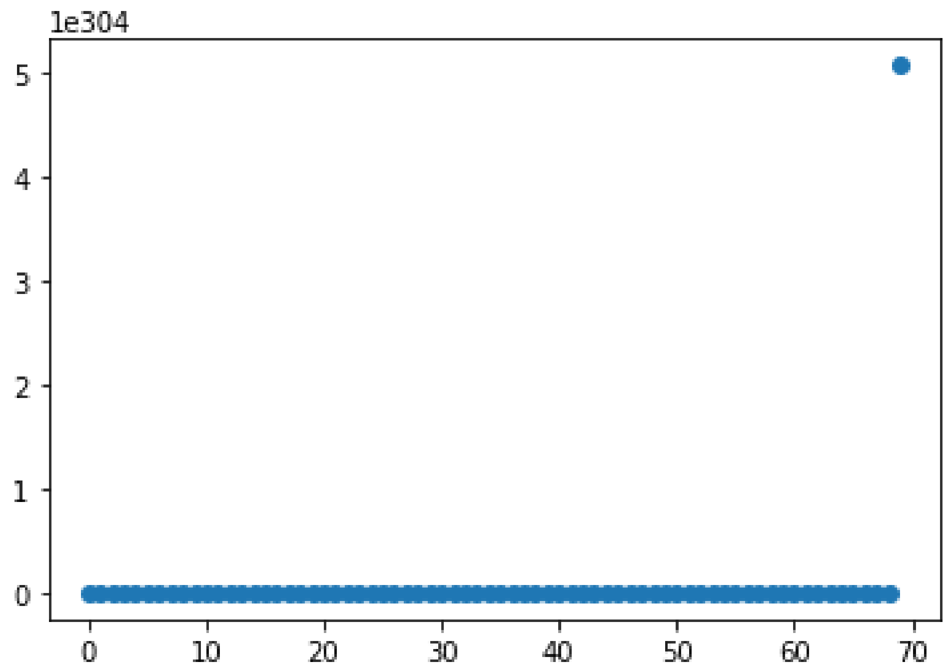
Using *graddescent* with $\eta = 0.05, t = 1000, \lambda = 0.05$, and plotting objective values.

With a now positive λ, we should see the penalty term add to the overall objective value, but then be countered by the gradient descent algorithm. We anticipate a curve that may increase but lower (or plateau) around some stable objective value.

Issue: Due to the aforementioned growing objective values, once again the plot below is invalid. Unfortunately, after several attempts, I was unable to locate and resolve the issue in the *graddescent* or *objective* functions.

```
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in multiply
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:
RuntimeWarning: invalid value encountered in matmul

/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: invalid value encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:7:
RuntimeWarning: invalid value encountered in subtract
  import sys
```

```
<matplotlib.collections.PathCollection at 0x7f9511e28d50>
```



Comparing our final B_t to $B^\star$ found by *sklearn.linear_model.Ridge,* as well as objective values for the same.

Note that our current observation is hindered by the NaN value for our final B_t iterate. The code for the comparison is still provided.

```
[nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan]
[-240.49345797  258.67047891   73.25271117  -65.00487414   -2.83514758
  134.7737705   -21.95736263 -421.67076899  500.41202241    6.40014714
  271.7112833   -20.74464182 -164.6833219    64.02144235   53.23744142
  -38.38824979   23.28897473  -49.73278524   -9.78418665]
```

```
Our iterate's obj. value: nan
Sklearn obj. value: 391749.79178054346
```

Using *graddescent* for many η, and comparing to *sklearn.linear_model.Ridge.*

The below code would compare (across $\eta = 0.1, 0.01, 0.001$) the minimum final iterates B_t found by our algorithm, to the *sklearn* values. Once again, the unresolved NaN issue prevents any conclusions.

```
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:
RuntimeWarning: overflow encountered in matmul

/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: invalid value encountered in matmul
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:2:
RuntimeWarning: invalid value encountered in matmul

/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:4:
RuntimeWarning: overflow encountered in multiply
  after removing the cwd from sys.path.
/home/anmol/miniconda3/lib/python3.7/site-packages/ipykernel_launcher.py:7:
RuntimeWarning: invalid value encountered in subtract
  import sys
```

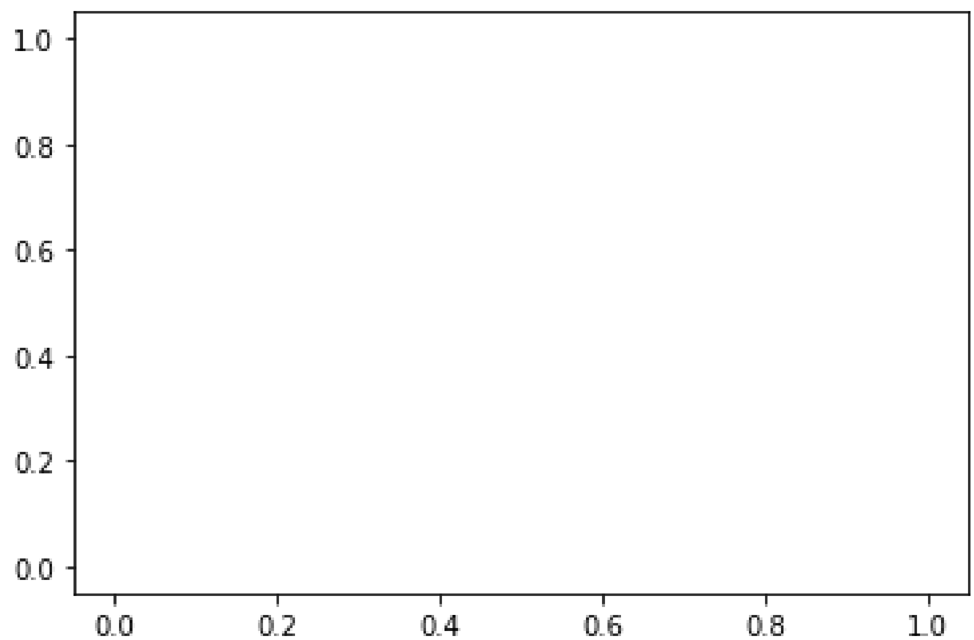
```
Objective value for our best final iterate: nan
Objective value for sklearn final iterate: 391749.79178054346
```

Modified ϵ-stationarity stopping criterion.

We notice that when trying different η values on a logarithmic scale, even our best final iterate B_T has objective value (slightly) higher than the sklearn function does, meaning the package function is consistently outperforming *e_graddescent*.

Note that the code is provided, but the NaN issue persists.

```
 [<matplotlib.lines.Line2D at 0x7f9513b37d90>]
```




```
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[-240.49345797  258.67047891   73.25271117  -65.00487414   -2.83514758
  134.7737705   -21.95736263 -421.67076899  500.41202241    6.40014714
  271.7112833   -20.74464182 -164.6833219    64.02144235   53.23744142
  -38.38824979   23.28897473  -49.73278524   -9.78418665]

Our iterate's obj. value: 441749.87561085785
Sklearn obj. value: 391749.79178054346

Objective value for our best (final) Beta iterate: 441749.87561085785
Objective value for sklearn iterate: 391749.79178054346
```

2 Exercise 2

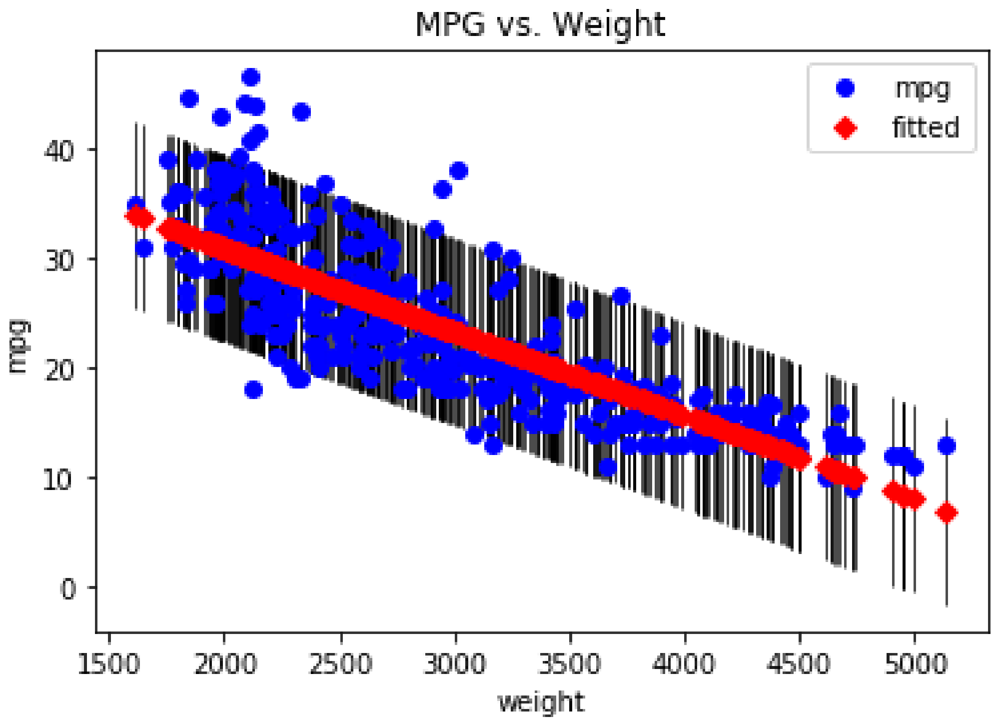
(a) Reading data.

(b) OLS with regards to MPG, weight.

The summary indicates that there is a relationship between the predictor and response, as the coefficient for weight is non-0, and the p-value suggests this is a statistically significant result. The relationship is negative, but also weak – as evidenced by the -0.0076 coefficient for weight – suggesting that as weight increases, MPG decreases slightly. However, the units for the response (mpg) are smaller than the units for the predictor (weight), which is measured in the thousands. So, a “unit” increase in weight may have little affect on MPG, but moving from the weight of one car to another may see large changes.

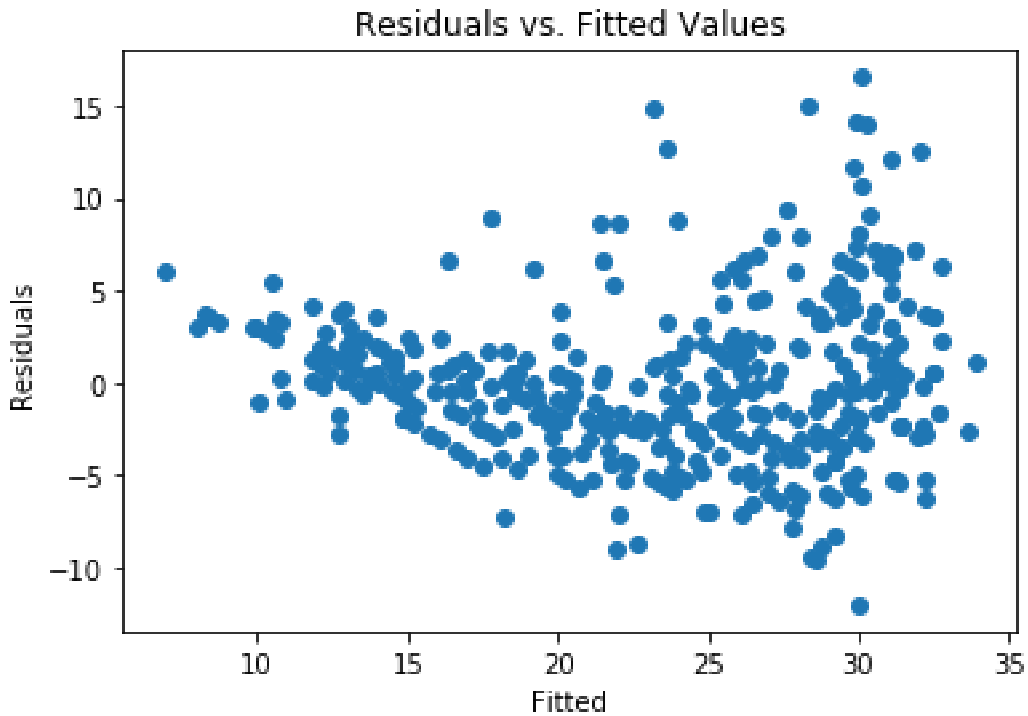
```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.693
Model:                            OLS   Adj. R-squared:                  0.692
Method:                 Least Squares   F-statistic:                     878.8
Date:                Fri, 17 Apr 2020   Prob (F-statistic):          6.02e-102
Time:                        23:47:32   Log-Likelihood:                -1130.0
No. Observations:                 392   AIC:                             2264.
Df Residuals:                     390   BIC:                             2272.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         46.2165      0.799     57.867      0.000      44.646      47.787
weight        -0.0076      0.000    -29.645      0.000      -0.008      -0.007
==============================================================================
Omnibus:                       41.682   Durbin-Watson:                   0.808
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               60.039
Skew:                           0.727   Prob(JB):                     9.18e-14
Kurtosis:                       4.251   Cond. No.                     1.13e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.13e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

(c) Plotting response vs. predictor.



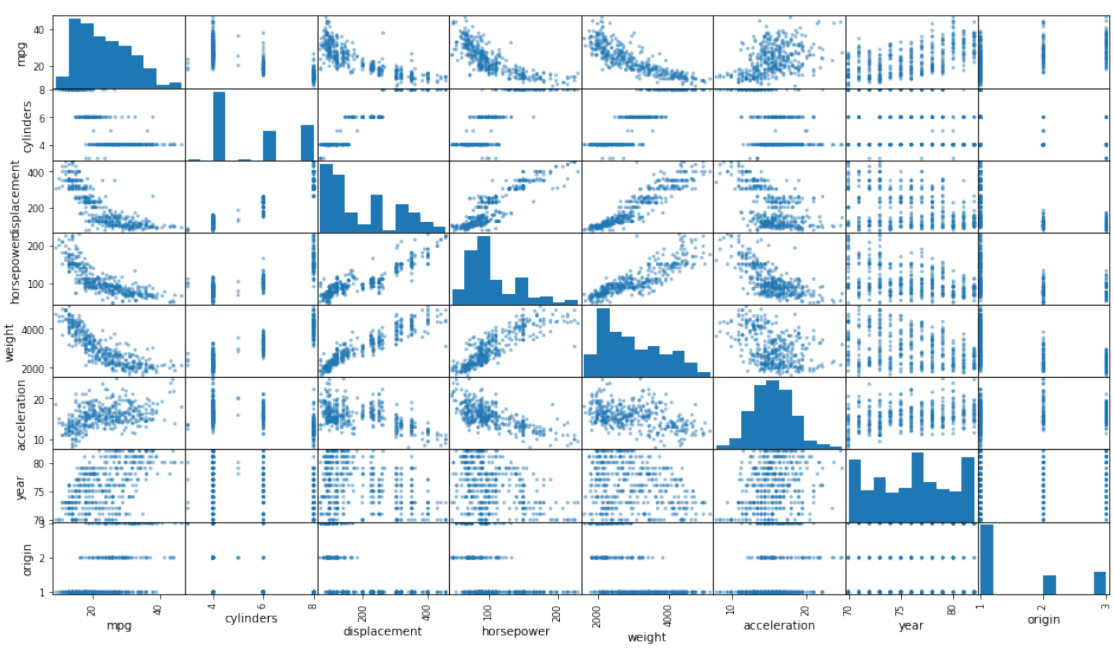
(d) Plotting residuals vs. fitted values.

The residuals have greater variation, and are farther from 0, as the fitted values increase. Essentially, for large fitted MPG values, the error increases. This could be because linear regression necessarily drives the fitted MPG values downwards as weight increases. This may be sensible for a while, but at a large enough weight, the actual change in MPG may not be best represented by a rigid line.



3 Exercise 3

(a) Scatter-matrix.



(b) Correlation matrix.

```
                   mpg  cylinders  displacement  horsepower    weight  \
mpg           1.000000  -0.777618     -0.805127   -0.778427 -0.832244
cylinders    -0.777618   1.000000      0.950823    0.842983  0.897527
displacement -0.805127   0.950823      1.000000    0.897257  0.932994
horsepower   -0.778427   0.842983      0.897257    1.000000  0.864538
weight       -0.832244   0.897527      0.932994    0.864538  1.000000
acceleration  0.423329  -0.504683     -0.543800   -0.689196 -0.416839
year          0.580541  -0.345647     -0.369855   -0.416361 -0.309120
origin        0.565209  -0.568932     -0.614535   -0.455171 -0.585005

              acceleration      year    origin
mpg               0.423329  0.580541  0.565209
cylinders        -0.504683 -0.345647 -0.568932
displacement     -0.543800 -0.369855 -0.614535
horsepower       -0.689196 -0.416361 -0.455171
weight           -0.416839 -0.309120 -0.585005
acceleration      1.000000  0.290316  0.212746
year              0.290316  1.000000  0.181528
origin            0.212746  0.181528  1.000000
```

(c) Multiple linear regression with MPG as the response, and everything else (bar 'name') as predictors.

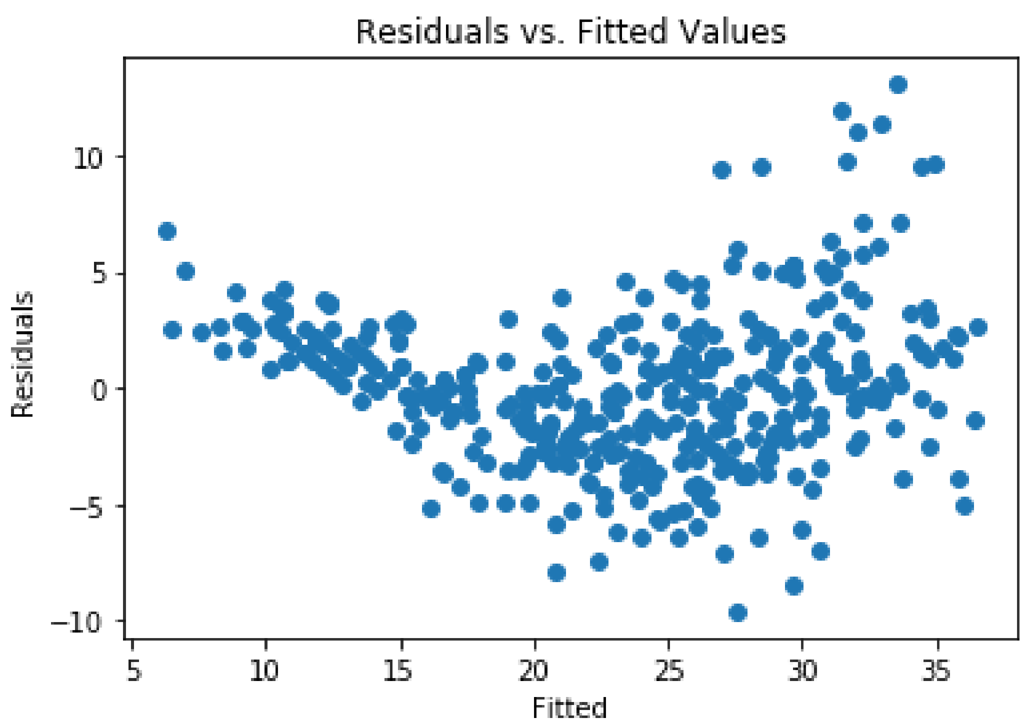
Some predictors certainly seem to have a relationship with the response. Those with a statistically significant relationship (by p-value) are: displacement, weight, year, and origin. The coefficient for year is 0.75, with a negligible p-value. This means MPG is notably correlated with year: unit increases in year cause an average increase of 0.75 in MPG.

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.821
Model:                            OLS   Adj. R-squared:                  0.818
Method:                 Least Squares   F-statistic:                     252.4
Date:                Fri, 17 Apr 2020   Prob (F-statistic):          2.04e-139
Time:                        23:47:39   Log-Likelihood:                -1023.5
No. Observations:                 392   AIC:                             2063.
Df Residuals:                     384   BIC:                             2095.
Df Model:                           7
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
```

```
const          -17.2184      4.644     -3.707      0.000     -26.350      -8.087
cylinders       -0.4934      0.323     -1.526      0.128      -1.129       0.142
displacement     0.0199      0.008      2.647      0.008       0.005       0.035
horsepower      -0.0170      0.014     -1.230      0.220      -0.044       0.010
weight          -0.0065      0.001     -9.929      0.000      -0.008      -0.005
acceleration     0.0806      0.099      0.815      0.415      -0.114       0.275
year             0.7508      0.051     14.729      0.000       0.651       0.851
origin           1.4261      0.278      5.127      0.000       0.879       1.973
==============================================================================
Omnibus:                       31.906   Durbin-Watson:                   1.309
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               53.100
Skew:                           0.529   Prob(JB):                     2.95e-12
Kurtosis:                       4.460   Cond. No.                     8.59e+04
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 8.59e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

(d) Plotting residuals vs. fitted values.

Again, we have a problem with the fit. As the predicted values get larger, so do the residuals, which implies the model is bad at predicting high MPG values, within reasonable margins of error.



(e) R-style interaction models.

The first model (above) has MPG as response, and solely the weight-acceleration interaction term as a predictor. The fitted coefficient (-0.0004) indicates a weakly negative relationship, but is statistically significant with p-value far below 0.05.

The second model again has MPG as response, and 3 predictors: the weight-acceleration interaction term, but weight and acceleration, too. The coefficients imply increasing weight has a (weakly) negative effect on MPG, while acceleration has a fairly positive effect. However, the acceleration term is not significant (p-value > 0.05), even though the interaction and weight terms are. This calls into question the validity of the entire model.

```
<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.341
Model:                            OLS   Adj. R-squared:                  0.339
Method:                 Least Squares   F-statistic:                     201.6
Date:                Fri, 17 Apr 2020   Prob (F-statistic):           3.57e-37
Time:                        23:47:39   Log-Likelihood:                -1279.5
No. Observations:                 392   AIC:                             2563.
Df Residuals:                     390   BIC:                             2571.
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
=======
                          coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-------
Intercept              40.5318      1.245     32.546      0.000      38.083
42.980
weight:acceleration    -0.0004   2.66e-05    -14.198      0.000      -0.000
-0.000
==============================================================================
Omnibus:                       32.873   Durbin-Watson:                   0.671
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               43.407
Skew:                           0.636   Prob(JB):                     3.75e-10
Kurtosis:                       4.019   Cond. No.                     1.82e+05
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.82e+05. This might indicate that there are
```

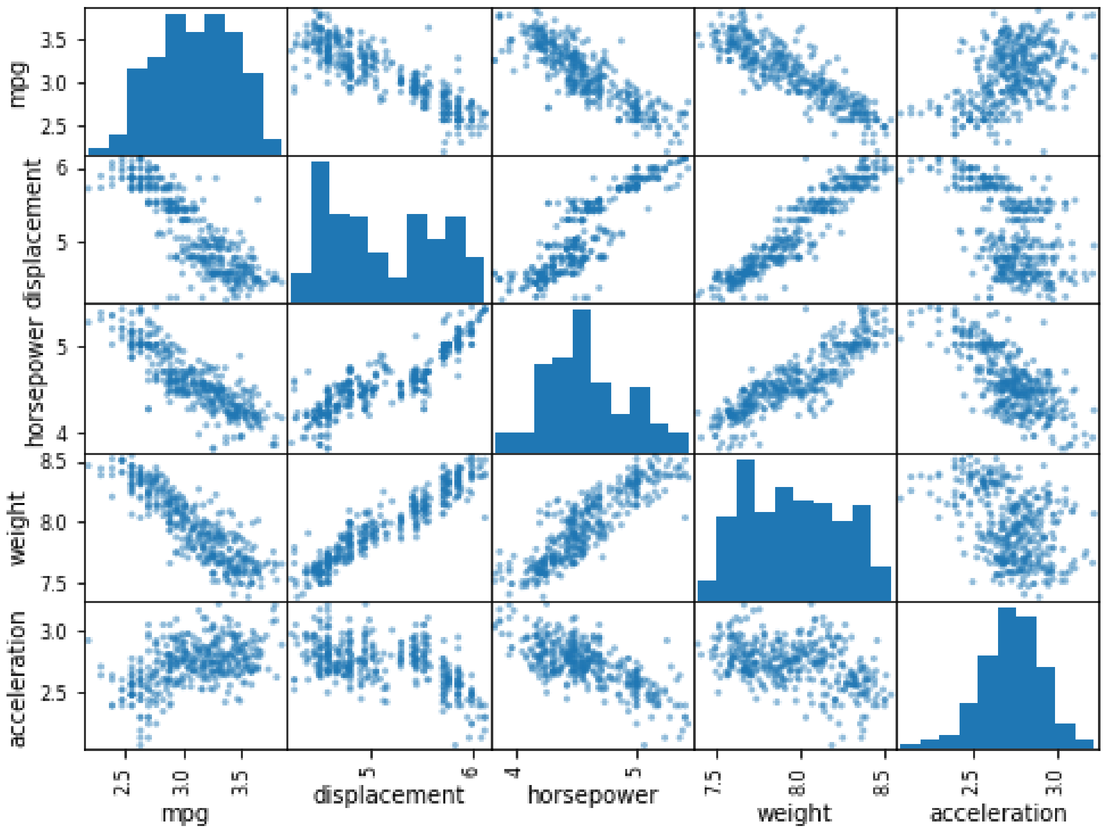
```
strong multicollinearity or other numerical problems.
"""

<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.706
Model:                            OLS   Adj. R-squared:                  0.704
Method:                 Least Squares   F-statistic:                     310.5
Date:                Fri, 17 Apr 2020   Prob (F-statistic):          9.83e-103
Time:                        23:47:39   Log-Likelihood:                -1121.3
No. Observations:                 392   AIC:                             2251.
Df Residuals:                     388   BIC:                             2266.
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
=======
                          coef    std err          t      P>|t|      [0.025
0.975]
--------------------------------------------------------------------------------
-------
Intercept              28.1398      4.872      5.776      0.000      18.562
37.718
weight                 -0.0032      0.001     -2.168      0.031      -0.006
-0.000
acceleration            1.1174      0.310      3.608      0.000       0.508
1.726
weight:acceleration    -0.0003   9.69e-05     -2.875      0.004      -0.000
-8.81e-05
==============================================================================
Omnibus:                       27.010   Durbin-Watson:                   0.809
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               34.785
Skew:                           0.555   Prob(JB):                     2.80e-08
Kurtosis:                       3.948   Cond. No.                     1.07e+06
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 1.07e+06. This might indicate that there are
strong multicollinearity or other numerical problems.
"""
```

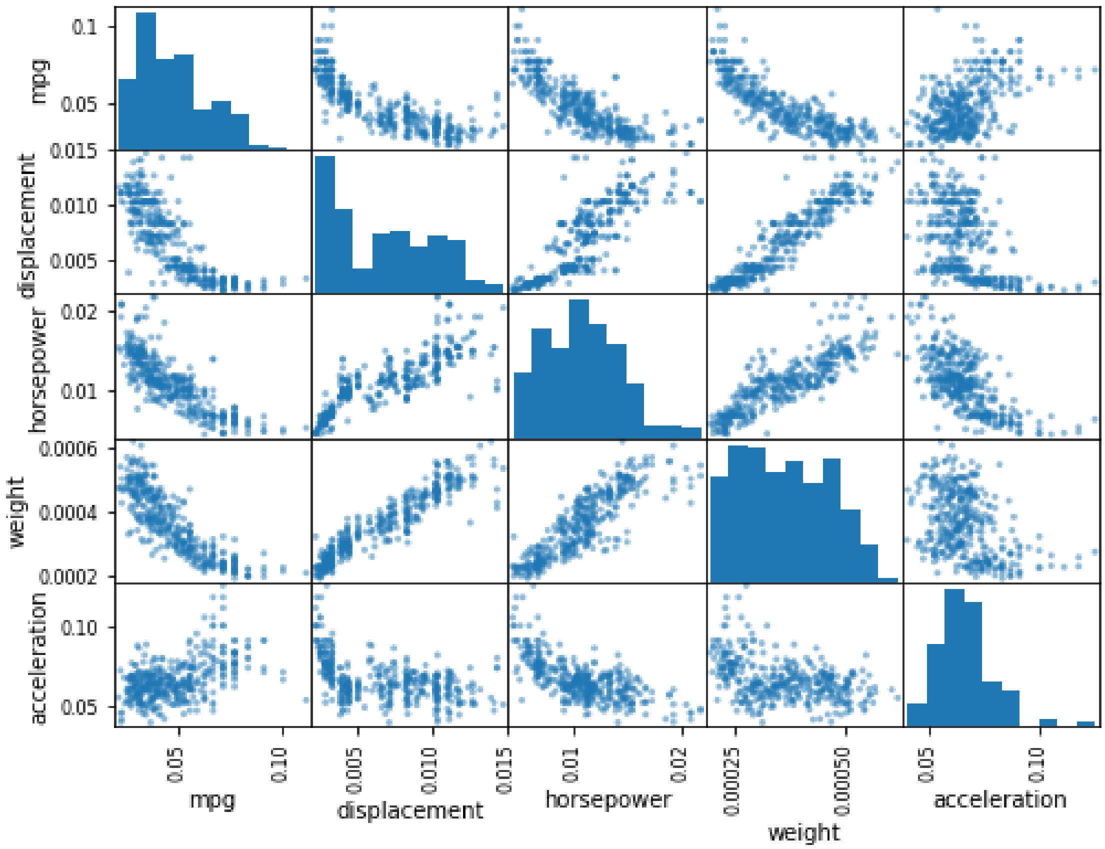
(f) Scatter-plots, now on *transformed* variables.

For simplicity, we observe only the continuous variables.

Our first transformation is log(X) on all variables; our second is 1/X on the same. Plots in the log(X) transform are largely unchanged. Some relationships in the original are sharpened(the transformed acceleration-horsepower points are gathered in a more compact negative trend). Still, no relationship is made clearer or fundamentally different.

Meanwhile, some relations in the 1/X transform are changed. For instance, the correlation between acceleration and displacement/weight/horsepower is now seemingly hyperbolic. This, of course, is simply a product of the reciprocal transformation that was applied. Any linear correlations will necessary become hyperbolic when inverted, so this transformation may not reveal further truth about the relationships.




mpg
displacement
horsepower
weight
acceleration
0.1
0.05
0.015
0.010
0.005
0.02
0.01
0.0006
0.0004
0.0002
0.10
0.05
0.10
0.005
0.010
0.015
0.00025
0.00050