Homework 1 - Math 4803

Dennis Goldenberg

January 2023

1 Theoretical Problems

1. (Finding LSS)

Using row reduction, we can see that:

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 3 & 6 \end{array}\right) \xrightarrow[R_2=R_2-R_1]{R_3=R_3-R_1} \left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 6 \end{array}\right)$$

Note here that $x_2 = 0$ from row 2 but $2x_2 = 6 \to x_2 = 3$ from row 3, making a perfect solution impossible. So, we will solve this system by solving the normal equation:

$$A^T b = A^T A x^* \to x^* = (A^T A)^{-1} A^T b$$

First, we compute $(A^T A)^{-1}$:

$$\left(\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} * \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}\right)^{-1}$$

$$= \left(\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}\right)^{-1}$$

Note that the determinant of this matrix is:

$$\det(A^T A) = 3(14) - 6(6) = 6$$

So, we find the inverse:

$$(A^T A)^{-1} = \frac{1}{6}\begin{pmatrix} 14 & -6 \\ -6 & 3 \end{pmatrix} = \begin{pmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{pmatrix}$$

Then, we multiply $A^T b$:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} * \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} = \begin{pmatrix} 6 \\ 18 \end{pmatrix}$$

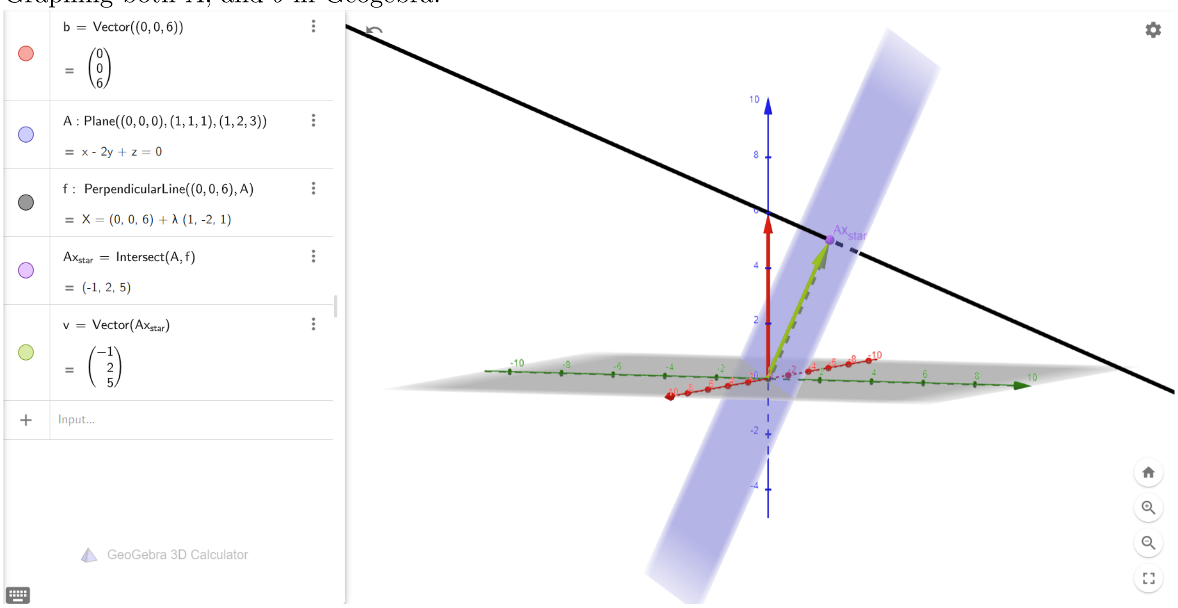
Then, we multiply $(A^T A)^{-1} A^T b$:

$$x^* = (A^T A)^{-1} A^T b = \begin{pmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{pmatrix} * \begin{pmatrix} 6 \\ 18 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \end{pmatrix}$$

Note that:

$$Ax^* = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} * \begin{pmatrix} -4 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}$$

Graphing both A, and b in Geogebra:



Here, A is the plane of the two columns going through the origin point. Note that the intersection point of the perpendicular line to b and the plane is $(-1, 2, 5)$, which is exactly Ax^*. Thus, Ax^* is the **orthogonal projection** of b onto the plane created by the columns of A going through the origin.

2. (Fitting function to $f(t) = c + p\sin(t) + q\cos(t)$)

a) We are given:

$$A_n \begin{bmatrix} c_n \\ p_n \\ q_n \end{bmatrix} = \overrightarrow{y}$$

For the the specific fitting outlined, we must structure A in the following way:

$$A_n = \begin{bmatrix} 1 & \sin(a_0) & \cos(a_0) \\ 1 & \sin(a_1) & \cos(a_1) \\ \vdots & \vdots & \vdots \\ 1 & \sin(a_n) & \cos(a_n) \end{bmatrix}$$

Then, to fit to the right set of points, we structure y the following way:

$$\overrightarrow{y} = \begin{bmatrix} g(a_0) \\ g(a_1) \\ \vdots \\ g(a_n) \end{bmatrix}$$

b) Let us compute $A_n^T A_n$:

$$A_n^T A_n = \begin{bmatrix} 1 & 1 & ... & 1 \\ \sin(a_0) & \sin(a_1) & ... & \sin(a_n) \\ \cos(a_0) & \cos(a_1) & ... & \cos(a_n) \end{bmatrix} * \begin{bmatrix} 1 & \sin(a_0) & \cos(a_0) \\ 1 & \sin(a_1) & \cos(a_1) \\ \vdots & \vdots & \vdots \\ 1 & \sin(a_n) & \cos(a_n) \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum_{i=0}^n \sin(a_i) & \sum_{i=0}^n \cos(a_i) \\ \sum_{i=0}^n \sin(a_i) & \sum_{i=0}^n \sin^2(a_i) & \sum_{i=0}^n \sin(a_i)\cos(a_i) \\ \sum_{i=0}^n \cos(a_i) & \sum_{i=0}^n \sin(a_i)\cos(a_i) & \sum_{i=0}^n \cos^2(a_i) \end{bmatrix}$$

Then, we multiply by a factor of $\frac{2\pi}{n}$:

$$\frac{2\pi}{n} A_n^T A_n = \begin{bmatrix} 2\pi & \frac{2\pi}{n}\sum_{i=0}^n \sin(a_i) & \frac{2\pi}{n}\sum_{i=0}^n \cos(a_i) \\ \frac{2\pi}{n}\sum_{i=0}^n \sin(a_i) & \frac{2\pi}{n}\sum_{i=0}^n \sin^2(a_i) & \frac{2\pi}{n}\sum_{i=0}^n \sin(a_i)\cos(a_i) \\ \frac{2\pi}{n}\sum_{i=0}^n \cos(a_i) & \frac{2\pi}{n}\sum_{i=0}^n \sin(a_i)\cos(a_i) & \frac{2\pi}{n}\sum_{i=0}^n \cos^2(a_i) \end{bmatrix}$$

To take the limit of each of these terms, we note that $\lim 2\pi = 2\pi$ so we don't have to worry about that. Taking the rest of these as Riemann integrals, we calculate each individually:

$$\int_0^{2\pi} \sin(a)da = -\cos(a)|_0^{2\pi} = 1 - 1 = 0$$

$$\int_0^{2\pi} \cos(a)da = \sin(a)|_0^{2\pi} = 0 - 0 = 0$$

For the next integral, we will use integration by parts and the trigonometric identity that $\sin^2(a) + \cos^2(a) = 1$:

$$\begin{aligned}
&\int_0^{2\pi} \sin^2(a)da = -\sin(a)\cos(a)|_0^{2\pi} + \int_0^{2\pi} \cos^2(a)da \\
&\rightarrow \int_0^{2\pi} \sin^2(a)da = 0 + \int_0^{2\pi} (1 - \sin^2(a))da \\
&\rightarrow \int_0^{2\pi} \sin^2(a)da = \int_0^{2\pi} 1da - \int_0^{2\pi} \sin^2(a)da \\
&\rightarrow 2\int_0^{2\pi} \sin^2(a)da = 2\pi \\
&\rightarrow \int_0^{2\pi} \sin^2(a)da = \pi
\end{aligned}$$

Using the trigonometric identity again:

$$\begin{aligned}
\int_0^{2\pi} \cos^2(a)da = \int_0^{2\pi} (1 - \sin^2(a))da &= \int_0^{2\pi} 1da - \int_0^{2\pi} \sin^2(a) \\
&= 2\pi - \pi \\
&= \pi
\end{aligned}$$

For the final integral, we use u-substitution, where $u = \sin(a)$:

$$\int_0^{2\pi} \sin(a)\cos(a)da = \int_0^0 udu = 0$$

As a result:

$$\lim_{n\to\infty} \frac{2\pi}{n} A_n^T A_n = \begin{bmatrix} 2\pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix}$$

Next, we calculate $A_n^T \overrightarrow{y}$:

$$\begin{bmatrix} 1 & 1 & ... & 1 \\ \sin(a_0) & \sin(a_1) & ... & \sin(a_n) \\ \cos(a_0) & \cos(a_1) & ... & \cos(a_n) \end{bmatrix} * \begin{bmatrix} g(a_0) \\ g(a_1) \\ \vdots \\ g(a_n) \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n g(a_i) \\ \sum_{i=0}^n \sin(a_i)g(a_i) \\ \sum_{i=0}^n \cos(a_i)g(a_i) \end{bmatrix}$$

Then, multiplying by $\frac{2\pi}{n}$ to put it in the correct form and taking the limit to infinity, effectively creating a Riemann integral, we get:

$$\lim_{n\to\infty} \frac{2\pi}{n} A_n^T \overrightarrow{y} = \begin{bmatrix} \int_0^{2\pi} g(a)da \\ \int_0^{2\pi} \sin(a)g(a)da \\ \int_0^{2\pi} \cos(a)g(a)da \end{bmatrix}$$

c) If we multiply both sides of the original linear system by $\frac{2\pi}{n} A_n^T$ and take the limit as $n \to \infty$, we get the new system:

$$\begin{bmatrix} 2\pi & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \pi \end{bmatrix} * \begin{bmatrix} c_\infty \\ p_\infty \\ q_\infty \end{bmatrix} = \begin{bmatrix} \int_0^{2\pi} g(a)da \\ \int_0^{2\pi} \sin(a)g(a)da \\ \int_0^{2\pi} \cos(a)g(a)da \end{bmatrix}$$

A simple matrix multiplication and 1 step of algebra gets us our coefficients:

$$\begin{bmatrix} c_\infty \\ p_\infty \\ q_\infty \end{bmatrix} = \begin{bmatrix} \frac{1}{2\pi}\int_0^{2\pi} g(a)da \\ \frac{1}{\pi}\int_0^{2\pi} \sin(a)g(a)da \\ \frac{1}{\pi}\int_0^{2\pi} \cos(a)g(a)da \end{bmatrix}$$

d) No, it is not necessarily true that all functions will be fit perfectly to $g(t)$ by $f_\infty(t)$. Some functions are not periodic, and therefore cannot be fit with periodic components and a constant term ($g(t) = t$ is an example). However, if the function was periodic, than you could find a perfect $f_\infty(t) = g(t)$ fit, because you have oscillating periodic components.

3. Section 2.4 Exercise 7

a) The test point is $X_1 = X_2 = X_3 = 0$. Now, we must compute the ℓ_2 norm of the difference to find euclidean distance. We calculate, in order of listed observation:

$$\left\| \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(0-0)^2 + (3-0)^2 + (0-0)^2} = 3$$

$$\left\| \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(2-0)^2 + (0-0)^2 + (0-0)^2} = 2$$

$$\left\| \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(0-0)^2 + (1-0)^2 + (3-0)^2} = \sqrt{10}$$

$$\left\| \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(0-0)^2 + (1-0)^2 + (2-0)^2} = \sqrt{5}$$

$$\left\| \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(-1-0)^2 + (0-0)^2 + (1-0)^2} = \sqrt{2}$$

$$\left\| \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\|_2 = \sqrt{(1-0)^2 + (1-0)^2 + (1-0)^2} = \sqrt{3}$$

b) The K-nearest neighbors classifier formula is as follows:

$$\mathbb{P}(y = j | X = x_0) = \frac{1}{K} \sum_{i \in N_0} 1_j(Y_i = j)$$

Here, N_0 is the set of K-nearest neighbors, and 1_j is the indicator function. When $K = 1$, we are only dealing with the nearest neighbor. As indicated by the Euclidean distances, the nearest neighbor to $X_1 = X_2 = X_3 = 0$ is observation 5. Note that $Y_5 =$ green. Let us call our test point y_t. So:

$$\mathbb{P}(y_t = \text{green} | X_1 = X_2 = X_3 = 0) = 1$$

Thus, our prediction is that the test point is **green**, as this is higher than 0.5, which was the defined decision boundary.

c) Since $K = 3$, we must now find the 3 closest neighbors to y_t. By euclidean distance, these are Observation 5, Observation 6, and Observation 2. Observation 5, as previously stated, was observed green, but Observations 2 and 6 were observed red. Therefore:

$$\mathbb{P}(y_t = red | X_0 = X_1 = X_2 = 0) = \frac{2}{3} > 0.5$$

Thus, our prediction is that the test point is **red**.

d) For an extremely non-linear set, we would expect the best value of K, or the one closest to the bayes classifier, to be **small**. This is because, due to the non-linearity, we need a flexible deicision boundary, which is achievable with smaller K values.

4. Section 3.7 Exercise 3

a) **iv** is correct. Note that the equation is:

$$\widehat{salary}_i = 50{+}20GPA_i{+}.07IQ_i{+}35Level_i{+}.01(GPA_i{*}IQ_i){-}10(GPA_i{*}Level_i)$$

To find the marginal effect of *Level* on the salary, we take the derivative with respect to level to get:

$$\frac{d\widehat{salary}_i}{d(Level_i)} = 35 - 10(GPA_i)$$

Note that, for $GPA > 3.5$, this marginal effect is positive, meaning that, for college students with high enough GPA, graduating college (so having a *Level* = 1) actually increases their starting salary; so, iv is true.

b) Thus is a specific observation where $IQ_i = 110, GPA_i = 4.0, Level_i = 1$. So:

$$\begin{aligned}\widehat{salary}_i &= 50 + 20GPA_i + .07IQ_i + 35Level_i + .01(GPA_i * IQ_i) - 10(GPA_i * Level_i)\\ &= 50 + 20(4.0) + .07(110) + 35(1) + .01(4.0 * 110) - 10(4.0 * 1)\\ &= 137.1\end{aligned}$$

The predicted starting salary is $\$137,100$.

c) **False**. GPA and IQ, when multiplied together, will likely produce very large values. So, even though the coefficient is small, the impact in any given observation may still be significant. In addition, we don't know the standard error of the coefficient, so we cannot comment on its statistical significance.

5. Section 3.7 Exercise 4

a) The **RSS for the cubic regression would likely be lower than the RSS for the linear regression for the training dataset**. This is because the cubic regression takes more observation information into account and thus can fit to to the training data set more accurately. In the worst possible case, where a squared term or a cubic term have no bearing on the training data, their coefficients can be lowered to near zero to have a smaller effect, which would effectively approximate an almost-linear regression. Thus, the RSS for cubic regression would be at or lower the RSS for linear regression.

b) Since the true relationship is linear, the **RSS for the linear regression would likely be lower than the RSS for the cubic regression for the testing dataset**. Since the model is first trained on the training dataset, the cubic regression, with its extra and irrelevant terms, is likely overfit to the training dataset. THerefore, it will be less accurate as it is less generalizable, and thus have a higher RSS than the linear regression, that is more indicative of the actual data.

c) Once again, the **RSS for the cubic regression would likely be lower than the RSS for linear regression in the training dataset**. Despite the relationship between X and Y changing from part a), what has not changed is the fact that the features in the linear regression are just a subset of the features in the cubic regression. Thus, the cubic regression would have more features that and thus be more flexible, making the residual sum of squares at least as good, if not lower.

d) **We do not have enough information to tell which RSS would be smaller**. Because the non-linearity between X and Y is not specified, we cannot conclude which model, cubic or linear, is more generalizable to this relationship. For this reason, we cannot make an educated guess on this question.

6. Section 3.7 Exercise 5

Proof. We can substitute the formula for $\hat{\beta}$ into the formula and algebraically manipulating:

$$\begin{aligned}
\hat{y}_i &= x_i\hat{\beta} \\
&= x_i * \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i'=1} x_{i'}^2} \\
&= y_1 \left(\frac{x_1 x_i}{\sum_{i'=1} x_{i'}^2} \right) + y_2 \left(\frac{x_2 x_i}{\sum_{i'=1} x_{i'}^2} \right) + ... + y_n \left(\frac{x_n x_i}{\sum_{i'=1} x_{i'}^2} \right) \\
&= a_1 y_1 + a_2 y_2 + ... + a_n y_n \\
&= \sum_{i'=1}^{n} a_{i'} y_{i'}
\end{aligned}$$

Here:

$$a_{i'} = \frac{x_{i'} x_i}{\sum_{i'=1} x_{i'}^2}$$

□

7. Section 3.7 Exercise 6

Proof. The simple linear regression model is as follows:

$$y_i = \beta_0 + \beta_1 x_i$$

Given a sample, equation 3.4 in the textbook gives us the estimators for the coefficients that minimize the residual sum of squares:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x}
\end{aligned}$$

Note that, if we plug in $\bar{x}$ for our explanatory variable, using these estimators for β_0 and β_1 gives us:

$$\begin{aligned}
y_i &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
&= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \\
&= \bar{y}
\end{aligned}$$

So, the least squares line indeed passes through $(\bar{x}, \bar{y})$. $\square$

2 Programming Problems

Programming Problem 1

1) Done in `HW1_coding.py`

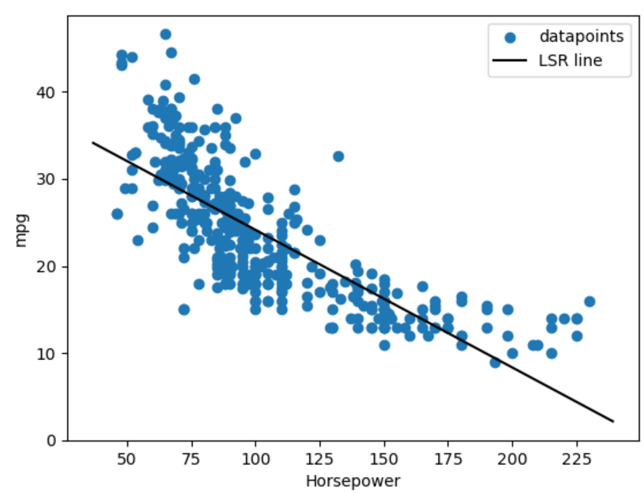
2) After running the regression (and removing the rows of data with unknown horsepower) we obtain the following regression:

```
==============================================================================
Dep. Variable:                    mpg   R-squared:                       0.606
Model:                            OLS   Adj. R-squared:                  0.605
Method:                 Least Squares   F-statistic:                     599.7
Date:                Wed, 25 Jan 2023   Prob (F-statistic):           7.03e-81
Time:                        13:54:46   Log-Likelihood:                -1178.7
No. Observations:                 392   AIC:                             2361.
Df Residuals:                     390   BIC:                             2369.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     39.9359      0.717     55.660      0.000      38.525      41.347
horsepower    -0.1578      0.006    -24.489      0.000      -0.171      -0.145
==============================================================================
Omnibus:                       16.432   Durbin-Watson:                   0.920
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               17.305
Skew:                           0.492   Prob(JB):                     0.000175
Kurtosis:                       3.299   Cond. No.                         322.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Sum of squared Residuals:  9385.915871932419
```

So, $\widehat{mpg} = 39.9359 - .1578(horsepower)$.

3) After creating the graph, the plot looked as follows:



4) The regression line **does not seem to fit the data well** because the relationship looks non-linear (likely exponential decay), and the RSS is high, as $RSS = 9385.916$.

Programming Problem 2

1) The correlation is as follows:

	mpg	cylinders	displacement	horsepower	weight	acceleration	year	origin
mpg	1.000000	-0.777618	-0.805127	-0.778427	-0.832244	0.423329	0.580541	0.565209
cylinders	-0.777618	1.000000	0.950823	0.842983	0.897527	-0.504683	-0.345647	-0.568932
displacement	-0.805127	0.950823	1.000000	0.897257	0.932994	-0.543800	-0.369855	-0.614535
horsepower	-0.778427	0.842983	0.897257	1.000000	0.864538	-0.689196	-0.416361	-0.455171
weight	-0.832244	0.897527	0.932994	0.864538	1.000000	-0.416839	-0.309120	-0.585005
acceleration	0.423329	-0.504683	-0.543800	-0.689196	-0.416839	1.000000	0.290316	0.212746
year	0.580541	-0.345647	-0.369855	-0.416361	-0.309120	0.290316	1.000000	0.181528
origin	0.565209	-0.568932	-0.614535	-0.455171	-0.585005	0.212746	0.181528	1.000000

The variables that appear highly correlated (having an absolute value of correlation above .8) are:

- mpg and displacement; mpg and weight
- cylinders and displacement, cylinders and horsepower, cylinders and weight
- displacement and weight
- horsepower and weight

2) After running the regression, we get:

```
                            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:                Thu, 26 Jan 2023   Prob (F-statistic):          2.04e-139
Time:                        16:56:48   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
-------------------------------------------------------------------------------
Intercept    -17.2184      4.644     -3.707      0.000     -26.350      -8.08
cylinders     -0.4934      0.323     -1.526      0.128      -1.129       0.14
displacement   0.0199      0.008      2.647      0.008       0.005       0.03
horsepower    -0.0170      0.014     -1.230      0.220      -0.044       0.01
weight        -0.0065      0.001     -9.929      0.000      -0.008      -0.00
acceleration   0.0806      0.099      0.815      0.415      -0.114       0.27
year           0.7508      0.051     14.729      0.000       0.651       0.85
origin         1.4261      0.278      5.127      0.000       0.879       1.97
==============================================================================
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
==============================================================================
```

The residual sum of squares is:

```
Sum of squared Residuals:  4252.212530440175
```

3) The multuple regression seems better, as it has a lower SSR and higher R-squared. However, the high correlation between predictor variables may leave some coefficients biased. This leads me to deduce that the simple linear regression is actually better.

Programming Problem 3

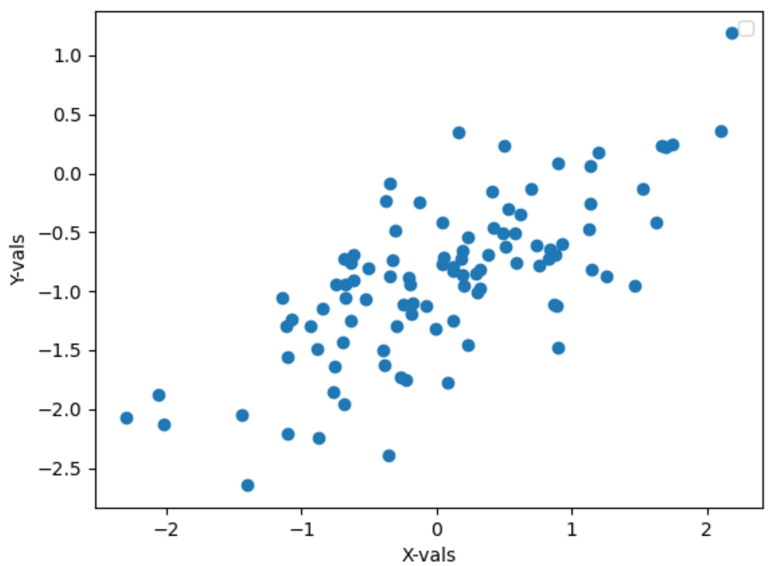
a) Done in `HW1_coding.py`

b) Done in `HW1_coding.py`

c) The resultant Y vector had a length of 100; here $\beta_0 = -1$ and $\beta_1 = .5$. Here was the Y vector generated:

```
Y:  [-0.4113916  -0.69362435 -1.06234006 -1.23969505 -1.11475211 -2.06607813
  0.24268411 -1.85745375 -0.9735897  -1.10837791 -0.95550469 -1.87249066
 -0.73812828 -1.62178515 -0.25784229 -2.20608734 -1.10556186 -2.24681539
 -0.41818427 -0.50414212 -1.56261807 -0.81521895  0.08767333  0.23479804
 -1.47856296 -0.72378191 -0.24761974 -1.29887887 -1.73357806 -0.30314961
 -1.43629053 -1.50033708 -1.95861542 -1.14733407 -0.93921963 -1.31809766
 -1.29836701 -1.45496285  0.2308316  -0.60569427 -1.18920266 -1.49468742
 -0.93913607  0.22143312 -0.70986346 -1.24964722 -0.86563169  0.3593177
 -0.82367324 -0.35012274 -1.00497323 -2.39354381 -1.0518468  -0.08118154
 -0.88376489 -0.75676602 -0.64873067 -0.59397605 -0.84850163 -1.11843878
 -1.6357462  -0.87207934 -0.61913551 -1.29736699 -0.50813526 -1.12513744
  0.05898229 -0.13332464  1.18813757 -2.64642863 -2.04551525 -0.80148949
  0.34418139 -0.68623293 -0.82034803 -2.12425773 -0.48737345 -0.72966661
 -0.54491771 -0.77889521 -1.74744345 -0.94360517 -0.6551269  -0.14836124
 -0.95607365 -1.24917671 -1.05395059 -0.69084956 -0.79875683 -0.4718144
  0.17962822 -0.72267543 -0.23531312 -0.76383685 -0.45872792 -1.77504914
 -0.8707672  -0.76806047 -0.90452459 -0.12876294]
```

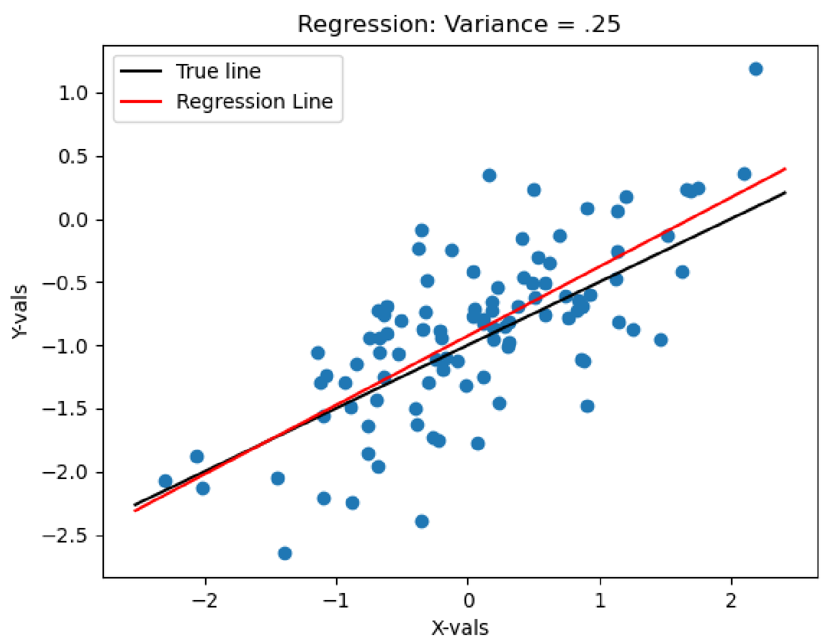
d) The following scatterplot was generated:



There does appear a strong correlation and a linear one at that.

e) The regression produced the following:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.522
Model:                            OLS   Adj. R-squared:                  0.517
Method:                 Least Squares   F-statistic:                     107.0
Date:                Thu, 26 Jan 2023   Prob (F-statistic):           2.20e-17
Time:                        19:02:11   Log-Likelihood:                -65.124
No. Observations:                 100   AIC:                             134.2
Df Residuals:                      98   BIC:                             139.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.9265      0.047    -19.717      0.000      -1.020      -0.833
xvars          0.5477      0.053     10.342      0.000       0.443       0.653
==============================================================================
Omnibus:                        0.898   Durbin-Watson:                   2.157
Prob(Omnibus):                  0.638   Jarque-Bera (JB):                0.561
Skew:                          -0.172   Prob(JB):                        0.755
Kurtosis:                       3.127   Cond. No.                         1.15
==============================================================================
```
Here $\widehat{\beta_0} = -.9265$ and $\widehat{\beta_1} = .5477$. They are fairly close to the true parameters in the model.
f) The population regression and least squares regression are shown here:



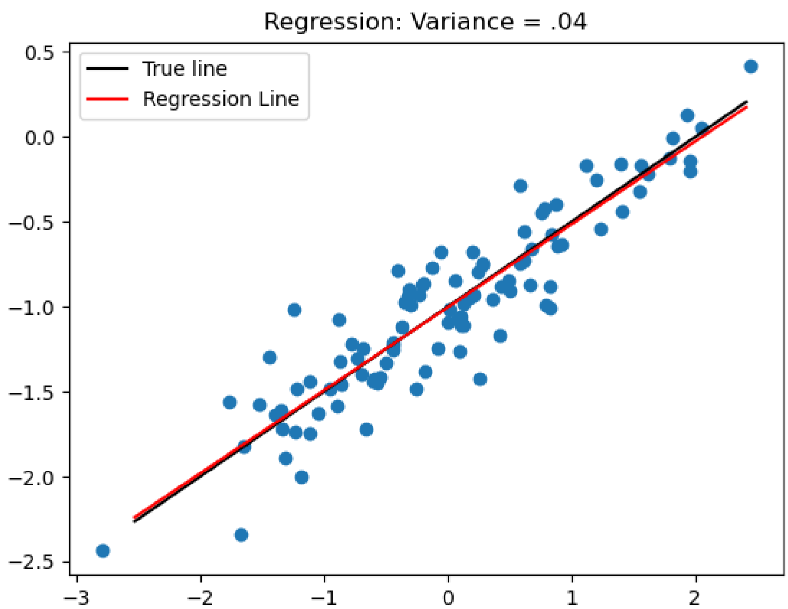
Here, the "true line" represents population regression, and the "regression line" represents the least squares regression.
g) After running the regression with x^2, we get:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.522
Model:                            OLS   Adj. R-squared:                  0.512
Method:                 Least Squares   F-statistic:                     52.96
Date:                Thu, 26 Jan 2023   Prob (F-statistic):           2.83e-16
Time:                        22:25:03   Log-Likelihood:                -65.107
No. Observations:                 100   AIC:                             136.2
Df Residuals:                      97   BIC:                             144.0
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.9325      0.058    -16.158      0.000      -1.047      -0.818
xvars          0.5468      0.053     10.229      0.000       0.441       0.653
xvarssq        0.0077      0.043      0.181      0.856      -0.077       0.092
==============================================================================
Omnibus:                        0.893   Durbin-Watson:                   2.152
Prob(Omnibus):                  0.640   Jarque-Bera (JB):                0.552
Skew:                          -0.170   Prob(JB):                        0.759
Kurtosis:                       3.132   Cond. No.                         2.10
==============================================================================
```

Note that there is very little statistical significance to the x^2 term, and the R^2 term does not change at all. This points to no improvement in fit with the inclusion of the x^2 term.

h) Here was the result with the standard deviation reduced to .2:

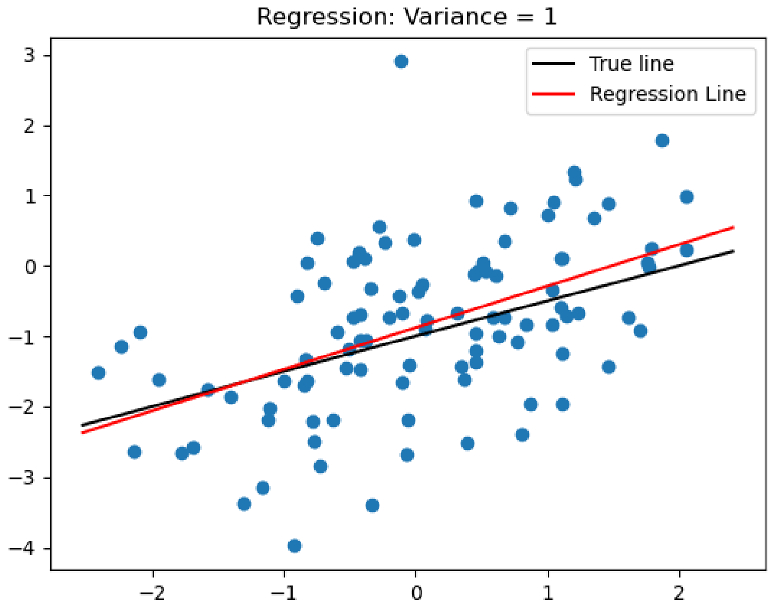
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.842
Model:                            OLS   Adj. R-squared:                  0.841
Method:                 Least Squares   F-statistic:                     523.3
Date:                Thu, 26 Jan 2023   Prob (F-statistic):           4.35e-41
Time:                        22:39:15   Log-Likelihood:                 12.946
No. Observations:                 100   AIC:                            -21.89
Df Residuals:                      98   BIC:                            -16.68
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -1.0040      0.021    -46.749      0.000      -1.047      -0.961
xvars          0.4888      0.021     22.876      0.000       0.446       0.531
==============================================================================
Omnibus:                        0.426   Durbin-Watson:                   1.977
Prob(Omnibus):                  0.808   Jarque-Bera (JB):                0.121
Skew:                          -0.045   Prob(JB):                        0.941
Kurtosis:                       3.145   Cond. No.                         1.01
==============================================================================
```



Note that $\widehat{\beta_0} = -1.004$ and $\widehat{\beta_1} = .4888$, which are much closer estimates to the true parameters, and the regression line is almost exactly overlapping the true line. This makes sense, as our deviation from said line was lowered.

i) With the standard deviation (and variance) increased to 1, we run the same regression with newly generated points:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.257
Model:                            OLS   Adj. R-squared:                  0.250
Method:                 Least Squares   F-statistic:                     33.96
Date:                Thu, 26 Jan 2023   Prob (F-statistic):           7.19e-08
Time:                        22:54:49   Log-Likelihood:                -145.66
No. Observations:                 100   AIC:                             295.3
Df Residuals:                      98   BIC:                             300.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.8802      0.105     -8.375      0.000      -1.089      -0.672
xvars          0.5903      0.101      5.828      0.000       0.389       0.791
==============================================================================
Omnibus:                        3.633   Durbin-Watson:                   1.972
Prob(Omnibus):                  0.163   Jarque-Bera (JB):                3.566
Skew:                           0.192   Prob(JB):                        0.168
Kurtosis:                       3.842   Cond. No.                         1.07
==============================================================================
```



Note that $\widehat{\beta_0} = -0.8802$ and $\widehat{\beta_1} = .5903$, which are further estimates to the true parameters then in the original $\sigma_\epsilon^2 = .25$ case as well as the reduced $\sigma_\epsilon^2 = .04$ case. This makes sense, as the points vary further away due to the increase in variance of error.