Homework 3: Multiple Linear Regression

Name: HARISH NANDHAN SHANMUGAM

This assignment is due on Gradescope by **Friday February 14th at 5:00PM**. If you submit the assignment by this deadline, you will receive 2 bonus points. If you need a little extra time, you may submit your work by **Monday February 17th at 5:00PM**. Your solutions to theoretical questions should be done in Markdown directly below the associated question. Your solutions to computational questions should include any specified R code and results as well as written commentary on your conclusions. Remember that you are encouraged to discuss the problems with your classmates, but **you must write all code and solutions on your own**.

NOTES:

- There are 2 total questions on this assignment.
- If you're not familiar with typesetting math directly into Markdown then by all means, do your work on paper first and then typeset it later. Remember that there is a reference guide linked here. **All** of your written commentary, justifications and mathematical work should be in Markdown.
- Because you can technically evaluate notebook cells in a non-linear order, it's a good idea to do Kernel → Restart & Run All as a check before submitting your solutions. That way if we need to run your code you will know that it will work as expected.
- It is **bad form** to make your reader interpret numerical output from your code. If a question asks you to compute some value from the data you should show your code output **AND** write a summary of the results in Markdown directly below your code.
- This probably goes without saying, but... For any question that asks you to calculate something, you **must show all work and justify your answers to receive credit**. Sparse or nonexistent work will receive sparse or nonexistent credit.

Problem 1 (50 Points)

PART A:

Prove that $t=\frac{\hat{\beta}_1}{\widehat{s.e.}\left(\hat{\beta}_1\right)}$ can be written in the form of a t-distributed random variable with $n-2$ degrees of freedom. Specifically, let Z be defined as $Z=\frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\acute{x}\right)^2}}}$, and $W=\frac{\left(n-2\right)\hat{\sigma}^2}{\sigma^2}$. Show that

\begin{align} |frac{Z}{|sqrt{W|big/(n-2)}} = |frac{|widehat|beta_1}{|widehat{s.e.}(|widehat| beta_1)} = t , |end{align}

which, by definition, means that $t \sim t(n-2)$.

$$\text{To prove: } \frac{z}{\sqrt{\frac{W}{n-2}}} = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} = t$$

Given:

$$z = \frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \acute{x})^2}}}$$

$$W = (n-2)\frac{\hat{\sigma}^2}{\sigma^2}$$

$$\frac{Z}{\sqrt{W}} = \frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \acute{x})^2}}}$$

$$¿ \frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \acute{x})^2}}} \times \sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2}}$$

$$¿ \frac{\hat{\beta}_1}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \acute{x})^2}}}$$

$$\frac{z}{\sqrt{W}} = \frac{\hat{\beta}_1}{\hat{\sigma}} \times \sqrt{\frac{\sum_{i=1}^{n}(x_i - \acute{x})^2}{(n-2)}}$$

$$\frac{z\sqrt{n-2}}{\sqrt{W}} = \frac{\hat{\beta}_1}{\hat{\sigma}} \times \sqrt{\sum_{i=1}^{n}(x_i - \acute{x})^2}$$

We know that the standard error of $\hat{\beta}_1$ is:

$$s.e\left(\hat{\beta}_1\right) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} \left(x_i - \acute{x}\right)^2}}$$

$$\therefore \frac{z}{\sqrt{\frac{W}{n-2}}} = \frac{\hat{\beta}_1}{s.e\left(\hat{\beta}_1\right)} = t$$

$$\text{(because } t = \frac{z}{\sqrt{\frac{v}{dof}}}\text{)}$$

where $n-2$ is the degrees of freedom.

PART B:

In class, we defined the *hat* or *projection* matrix as

$$H = X\left(X^T X\right)^{-1} X^T.$$

The goal of this question is to use the hat matrix to prove that the fitted values, $\hat{Y}$, and the residuals, $\hat{\varepsilon}$, are uncorrelated. We will do it in steps.

First, show that $\hat{Y} = HY$. That is, H "puts a hat on" Y.

Given HAT matrix:

$$H = X\left(X^T X\right)^{-1} X^T$$

To prove: $\hat{Y} = HY$

We know that: $Y = X\beta$ (Equation 1)

Multiply X^T on both sides:

$$X^T Y = X^T X \beta$$

Let's assume that $X^T X$ is invertible.

$$\left(X^T X\right)^{-1} X^T Y = \left(X^T X\right)^{-1} X^T X \beta$$

Since $\left(X^T X\right)^{-1} X^T X = I$

$$\left(X^T X\right)^{-1} X^T Y = I\beta$$

$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$ (Equation 2)

Substituting (2) into (1):

$$\hat{Y} = X\hat{\beta}$$

$$\hat{Y} = X\left(X^T X\right)^{-1} X^T Y$$

$$H = X\left(X^T X\right)^{-1} X^T \text{ (Given)}$$

$$\hat{Y} = H\,Y$$

$\therefore$ Hence proved.

PART C: Show that H is symmetric: $H = H^T$.

We know the transpose property:

$$\left(A\,B\right)^T = B^T A^T$$

$$\left(A^T\right)^T = A$$

$$H^T = \left(X\left(X^T X\right)^{-1} X^T\right)^T$$

$$¿\left(X^T\right)^T\left(\left(X^T X\right)^{-1}\right)^T X^T$$

Here, the inverse of a symmetric matrix is also symmetric:

$$\left(X^T X\right)^{-1} = \left(\left(X^T X\right)^{-1}\right)^T$$

So,

$$H^T = X\left(X^T X\right)^{-1} X^T$$

$$\therefore H^T = H$$

Hence proved.

PART D: Show that $H\left(I_n - H\right) = 0_n$, where 0_n is the zero matrix of size $n \times n$.

To prove: $H\left(I_n - H\right) = O_n$

$$H\left(I_n - H\right) = H\,I_n - H\,H \text{ (Equation 1)}$$

where I_n is the Identity matrix, so $H\,I_n = H$ (Equation 2)

$$H\,H = X\left(X^T X\right)^{-1} X^T X\left(X^T X\right)^{-1} X^T$$

We know that: $\left(X^T X\right)\left(X^T X\right)^{-1} = I$

So here,

$$H\,H = X\left(X^T X\right)^{-1} I\,X^T$$

$$H\,H = X\left(X^T X\right)^{-1} X^T$$

$$HH = H \text{ (Equation 3)}$$

Substituting (2) and (3) into (1):

$$H(I_n - H) = H - H$$

$$H(I_n - H) = 0_n$$

Hence proved.

PART E:

Stating that $\hat{Y}$ is uncorrelated with $\hat{\varepsilon}$ is equivalent to showing that these vectors are orthogonal. That is, we need to show that their dot product is zero:**

$$\hat{Y}^T \hat{\varepsilon} = 0.$$

Prove this result.

To prove: $\hat{Y}^T \hat{\varepsilon} = 0$

Error: $\hat{\varepsilon} = Y - \hat{Y}$

We know from Part B that: $\hat{Y} = HY$

$$\hat{\varepsilon} = Y - HY$$

$$\hat{\varepsilon} = (I - H)Y$$

$$(\hat{Y})^T = (HY)^T$$

$$¿ Y^T H^T$$

From Part C, we know: $H = H^T$

$$\therefore (\hat{Y})^T = Y^T H$$

$$\hat{Y}^T \hat{\varepsilon} = Y^T H (I - H) Y$$

From Part D, we know: $H(I_n - H) = O_n$

$$\therefore \hat{Y}^T \hat{\varepsilon} = Y^T O_n Y$$

$$\hat{Y}^T \hat{\varepsilon} = 0$$

Hence proved.

PART F: Why is this result important in the practical use of linear regression?

- The residuals are uncorrelated with the fitted values, so the Ordinary least square estimates will be remained unbiased which leads to the accurate interpretation of coefficients (orthogonal represents there is no relationship between feature and the response).

- The decomposition of SST = SSR + SSE is true only if it is orthogonal which makes the R^2 estimate a goodness of fit measure (model very well explains the variability in Y).
- The residuals and the fitted values orthogonal also shows that the model minimized error and the coefficients are accurate for the fitted line.

Problem 2 - Polynomial Regression via Multiple Linear Regression (50 points)

It's not too difficult to believe that some relationships between features and the response are nonlinear. Consider the following example, where the single feature x and the response y have a quadratic relationship of the form

$$Y = \frac{1}{4} - X + X^2 + \epsilon$$

Part A: Write code that samples $n = 30$ data points from the true model, fits the SLR model, and plots it against the data. It should be clear from both the picture and the R^2-value that the SLR model is a very poor fit of the data.

```
set.seed(16)
n = 30
x = seq(-2, 2, length.out = n)
e = rnorm(n)

y = (1/4) - x + x^2 + e
slr_model = lm(y ~ x)
summary(slr_model)

plot(x, y, main = "SLR Model Fitted on Quadratic Data",
     xlab = "X", ylab = "Y", col = "blue", pch = 16)
abline(slr_model, col = "red", lwd = 2)
```

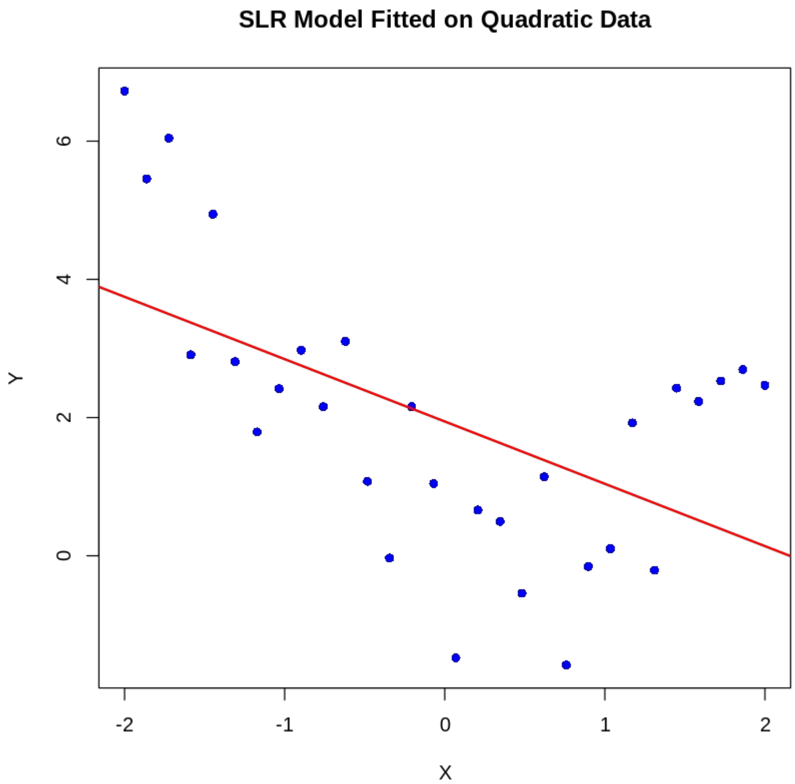
```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.3581 -1.1249 -0.3869  1.7150  2.9794 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1.9430     0.3176   6.118 1.33e-06 ***
x            -0.9020     0.2660  -3.391  0.00209 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.739 on 28 degrees of freedom
```

```
Multiple R-squared:  0.2911,	Adjusted R-squared:  0.2658 
F-statistic:  11.5 on 1 and 28 DF,  p-value: 0.002091
```



Part B: We can fit a **polynomial** model to the single-feature data be thinking of the polynomial features as features in a multiple linear regression model. If we make the association

$$x_1 = x \text{ and } x_2 = x^2$$

then we can fit a multiple linear regression of the form

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Fit this model and print the regression coefficients to the screen.

```
x1 = x
x2 = x^2
poly_model_mlr = lm(y~x1+x2)
summary(poly_model_mlr$coefficients)
summary(poly_model_mlr)
coef_poly_model = coef(poly_model_mlr)
coef_poly_model
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-0.9020 -0.2513  0.3994  0.1935  0.7412  1.0830 
```

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.9193 -0.6285  0.2057  0.7964  1.7267 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   0.3994     0.2778   1.438    0.162    
x1           -0.9020     0.1550  -5.821 3.40e-06 ***
x2            1.0830     0.1454   7.450 5.15e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.013 on 27 degrees of freedom
Multiple R-squared:  0.768,Adjusted R-squared:  0.7508 
F-statistic: 44.69 on 2 and 27 DF,  p-value: 2.714e-09

(Intercept)          x1          x2 
  0.3993549  -0.9020225   1.0830205 
```

Part C: Write down the estimated MLR model in terms of the features x_1 and x_2 as well as the interpretation of the associated polynomial model in terms of the single feature x. Does this model seem close to the true model that the data was generated from?

True Model

$$Y = 1/4 - X + X^2 + \epsilon$$

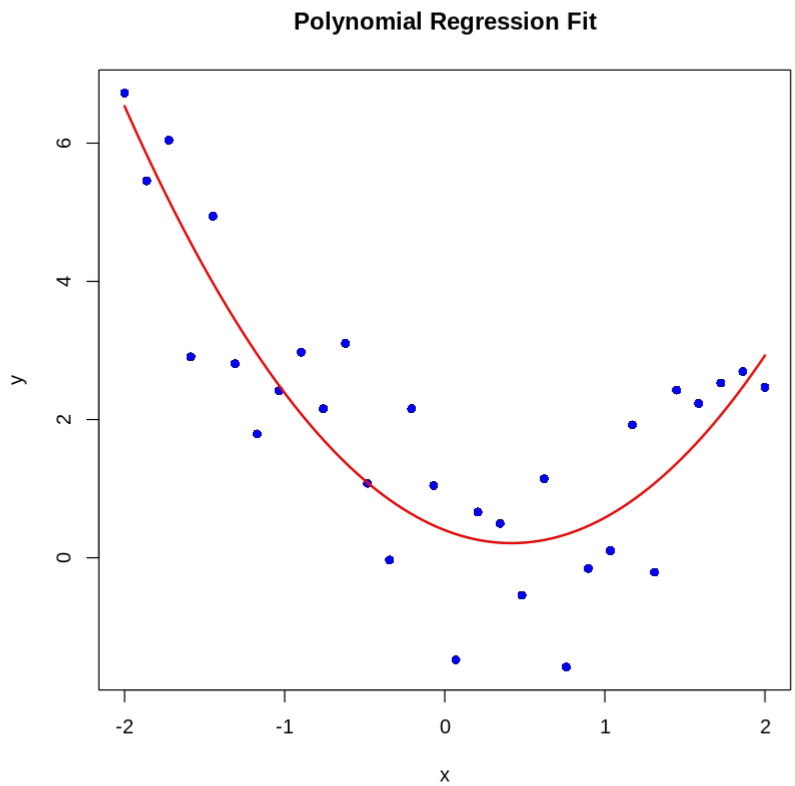
Estimated Multiple Linear Regression model

$$\hat{y} = 0.399 - 0.902\,x + 1.083\,x^2$$

Yes, the estimated multiple linear regression model is close to the true model the data was generated from because the $\widehat{\beta}_1$ is negative in both of the models, $\widehat{\beta}_2$ is positive in both the cases while there is a minor difference in numbers (1 vs 0.902) in $\widehat{\beta}_1$ and (1 vs 1.083) in $\widehat{\beta}_2$. And the intercept $\widehat{\beta}_0$ is positive in both of these cases although a small difference in value (0.25 vs 0.39). This MLR model is better than the SLR model.

Part D: Plot the obtained polynomial regression model against the data. Comment on the fit.

```r
plot(x, y, pch = 16, col = "blue", main = "Polynomial Regression Fit")
x_smooth = seq(min(x), max(x), length.out = 100)
y_pred = coef_poly_model[1] + coef_poly_model[2] * x_smooth +
coef_poly_model[3] * x_smooth^2
lines(x_smooth, y_pred, col = "red", lwd = 2)
```



- Here the polynomial ultiple linear regression model follows a U shaped curve.
- This curve captured the data points very well compared to the Simple linear regression line.
- The R squared score improved from 0.3 to 0.7 with this polynomial MLR model.
- This curve still has some residual variability does the model does not fully capture, since some of the points deviate significantly from the curve.

Problem

1.A) To prove $\frac{z}{\sqrt{W/(n-2)}} = \frac{\hat{\beta}_1}{\hat{s.e}(\hat{\beta}_1)} = t$

given, $z = \frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}$, $W = \frac{(n-2)\hat{\sigma}^2}{\sigma^2}$

$$\frac{z}{\sqrt{W}} = \frac{\frac{\hat{\beta}_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}}{\sqrt{\frac{(n-2)\hat{\sigma}^2}{\sigma^2}}} = \frac{\hat{\beta}_1}{\sqrt{\frac{\cancel{\sigma^2}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\sqrt{\frac{(n-2)\hat{\sigma}^2}{\cancel{\sigma^2}}}}$$

$$\frac{z(\sqrt{n-2})}{\sqrt{W}} = \frac{\hat{\beta}_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}$$

$$\frac{z}{\sqrt{\frac{W}{n-2}}} = \frac{\hat{\beta}_1}{\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}}$$

we know that, standard error of $\hat{\beta}_1$ is $\frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

$$\therefore \frac{z}{\sqrt{\frac{W}{n-2}}} = \frac{\hat{\beta}_1}{\hat{s.e}(\hat{\beta}_1)} = t \left(\text{because } t = \frac{z}{\sqrt{\frac{V}{\gamma}}}\right) \gamma\text{-degree of freedom}$$

---

2. given HAT matrix

$$H = X(X^TX)^{-1}X^T,$$

To prove, $\hat{Y} = HY$

we know that, $Y = X\beta \longrightarrow (1)$

Xply X^T on both sides

$$X^TY = X^TX\beta$$

lets Assume that X^TX is invertible

$$(X^TX)^{-1}X^TY = (X^TX)^{-1}X^TX\beta$$

we know that $A^{-1}A = I$

$$(X^TX)^{-1}X^TY = I\beta$$

$$\boxed{\vec{\beta} = (X^TX)^{-1}X^TY} \longrightarrow (2)$$

Substitute (2) in (1)

$$\vec{y} = X\vec{\beta}$$
$$\vec{y} = X(X^TX)^{-1}X^TY$$
$$H = X(X^TX)^{-1}X^T \text{ (given)}$$
$$\boxed{\vec{y} = HY}$$

$\therefore$ Hence proved

---

3. We know the transpose property

$$(AB)^T = B^TA^T$$
$$(A^T)^T = A$$
$$H^T = \left(X(X^TX)^{-1}X^T\right)^T$$
$$= (X^T)^T\left((X^TX)^{-1}\right)^T X^T$$

Here inverse of a symmetric matrix is also symmetric

i.e) $(X^TX)^{-1} = \left((X^TX)^{-1}\right)^T$

so,

$$H^T = X(X^TX)^{-1}X^T$$
$$\therefore \boxed{H^T = H}$$

Hence proved

---

4. To prove : $H(I_n - H) = 0_n$

$$H(I_n - H) = HI_n - HH \longrightarrow (1)$$

where I_n is the Identity matrix so $HI_n = H \longrightarrow (2)$

$$HH = X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T$$

we know that $(X^TX)(X^TX)^{-1} = I$.

so here,

$$HH = X(X^TX)^{-1}IX^T$$
$$HH = X(X^TX)^{-1}X^T$$
$$HH = H \longrightarrow (3)$$

substitute (2) & (3) in (1)

$$H(I_n - H) = H - H$$
$$\therefore \boxed{H(I_n - H) = 0_n}$$

Hence proved

---

5. To prove $\hat{y}^T\hat{e} = 0$

error $\hat{e} = Y - \hat{y}$

we know that from part B: $\hat{y} = HY$

$$\hat{e} = Y - HY$$
$$\boxed{\hat{e} = (I - H)Y}$$

$$(\hat{y})^T = (HY)^T$$
$$= Y^TH^T$$

here we know from part C: $H = H^T$

$$\boxed{(\hat{y})^T = Y^TH}$$

$$\hat{y}^T\hat{e} = Y^TH(I - H)Y$$

here we know from part D: $H(I_n - H) = 0_n$

so, $\hat{y}^T\hat{e} = Y^T0_nY$

$$\boxed{\hat{y}^T\hat{e} = 0}$$

Hence proved