PSTAT 126: Homework #3

1) Below (Before prob. 5)

2) We have been finding a set of optimal parameters for the linear least-squares regression minimization problem by identifying critical points, i.e. points at which the gradient of a function is the zero vector, of the following function:
$$F(\alpha_0, \ldots, \alpha_M) = \sum_{n=1}^{N}(\alpha_0 + \alpha_1 x_{n1} + \cdots + \alpha_M x_{nM} - y_n)^2 .$$
Help to justify this methodology in the following way. Letting $G: \mathbb{R}^d \to \mathbb{R}$ be any function that is differentiable everywhere, show that, if G has a local minimum at a point x_0, then its gradient is the zero vector there, i.e., $\nabla G(x_0) = \mathbf{0}$.

X_0 = local min.

$G(y) = G(X_0) + \nabla G(X_0)^T (y - X_0) + \text{h.o. term}$

$B(X_0, P)$ $\forall y \in B(X_0, P)$ $G(y) \geq G(X_0)$

$y = X_0 + \epsilon$ $\epsilon > 0$

$G(X_0 + \epsilon) - G(X_0) = \nabla G(X_0)^T(\epsilon) \geq 0$

$y = X_0 - \epsilon,$ $\epsilon > 0$

$G(X_0 - \epsilon) - G(X_0) = -\underbrace{\nabla G(X_0)^T}_{-1}(\epsilon) \geq \underset{-1}{\underline{0}}$

Since

$\nabla G(X_0)^T(\epsilon) \geq 0, \nabla G(X_0)^T(\epsilon) \leq 0$

$\nabla G(X_0) = 0$

3) A function $G: \mathbb{R}^d \to \mathbb{R}$ is called *convex* if it satisfies the following:

$$G(\lambda u + (1-\lambda)v) \leq \lambda G(u) + (1-\lambda)G(v),$$

for any $u, v \in \mathbb{R}^d$ and λ any scalar with $\lambda \in [0,1]$. Show that the function F in Question 2 above is convex using the following steps:

a) Show that the sum of convex functions is convex.

b) Define the function

$$H(\alpha_0, \ldots, \alpha_M) = \alpha_0 + \alpha_1 x_1 + \cdots + \alpha_M x_M - y,$$

for any scalar, fixed constants $x_1, \ldots, x_M, y$. Show that

$$H(\lambda u + (1-\lambda)v) = \lambda H(u) + (1-\lambda)H(v),$$

where u, v, and λ are as at the beginning of this question (Question 3) with $d = M + 1$.

c) Show that the function $f(t) = t^2$, t a scalar real number, is convex.

d) Explain how steps (a)-(c) can be put together to establish that the function F in Question 2 is convex.

$F: \mathbb{R}^d \to \mathbb{R}$ convex $\quad u, v \in \mathbb{R}^d$

$G: \mathbb{R}^d \to \mathbb{R}$ convex $\quad \lambda \in [0,1]$

$F+G: \mathbb{R}^d \to \mathbb{R}$ convex

a) $(F+G)(\lambda u + (1-\lambda)v)$

$= F(\lambda u + (1-\lambda)v) + G(\lambda u + (1-\lambda)v)$

$\lambda F(u) + (1-\lambda)F(v) + \lambda G(u) + (1-\lambda)G(v)$

F and G are convex

$= \lambda[F(u) + G(u)] + 1-\lambda[F(v) + G(v)]$

$= \lambda[(F+G)(u)] + 1-\lambda[(F+G)(v)]$

Sum of (F+G) is convex

b) $H(\alpha_0, \ldots, \alpha_m) = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_m x_m - y$ fixed constants: $x_1, \ldots x_M, y$

$H(\lambda u + (1-\lambda)v) = \lambda H(u) + (1-\lambda)H(v)$

$u = \alpha_0, \alpha_1, \ldots, \alpha_m$

$v = \beta_0, \beta_1, \ldots \beta_m$

$= H(\lambda(\alpha_0, \alpha_1, \ldots, \alpha_m) + (1-\lambda)(\beta_0, \beta_1, \ldots \beta_m))$

$= H(\lambda\alpha_0, \lambda\alpha_1, \ldots, \lambda\alpha_m) + ((1-\lambda\beta_0), (1-\lambda\beta_1), \ldots, (1-\lambda\beta_m))$

$= H(\lambda\alpha_0 + (1-\lambda)\beta_0, \lambda\alpha_1 + (1-\lambda)\beta_1, \ldots, \lambda\alpha_m + (1-\lambda)\beta_m)$

$= (\lambda\alpha_0 + (1-\lambda)\beta_0) + (\lambda\alpha_1 + (1-\lambda)\beta_1)x_1 + \ldots + (\lambda\alpha_m + (1-\lambda)\beta_m)x_m - y$

$= \lambda(\alpha_0 + \alpha_1 x_1 + \ldots + \alpha_m x_m - y) + (1-\lambda)(\beta_0 + \beta_1 x_m + \ldots + \beta_m x_m - y)$

$= \lambda H(u) + (1-\lambda)H(v)$

c) $f(t) = t^2$, t: scalar, convex

$f(\lambda u + (1-\lambda)v)$

$= (\lambda u + (1-\lambda)v)^2$

$= \lambda^2 u^2 + (1-\lambda)^2 v^2 + 2\lambda(1-\lambda)uv$

$= f(\lambda u + (1-\lambda)v) - \lambda f(u) - (1-\lambda)f(v)$

$= \lambda^2 u^2 + (1-\lambda)^2 v^2 + 2\lambda(1-\lambda)uv - \lambda u^2 - (1-\lambda)v^2$

$= -\lambda(1-\lambda)(u-v)^2 \leq 0$

f is a convex f'n

d) Function F is a convex f'n by first doing step b, by defining the function, then step C to show it is convex, and lastly step a to show the sum is convex, verifying the f'ns are as well.

4) Show that if a convex function $G: \mathbb{R}^d \to \mathbb{R}$ has a local minimum at a point x_0, then this local minimum is also a global minimum for the function over all of its domain $\mathbb{R}^d$.

proof by contradiction:

Let x_0 be a local min but not a global min.

$$y \in \mathbb{R}^d \quad G(y) < G(x_0) \quad \lambda \in [0,1]$$

$$G(\lambda y + (1-\lambda)x_0) \leq \lambda G(y) + (1-\lambda)G(x_0)$$

$$G(y) - G(x_0) < 0$$

$$\lambda G(y) + (1-\lambda)G(x_0) = \lambda(G(y) - G(x_0)) + G(x_0) < G(x_0)$$

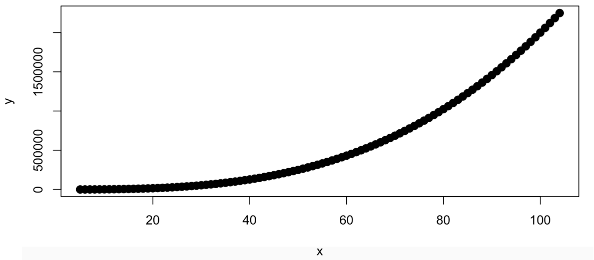
$$G(\lambda(y) + (1-\lambda)x_0) < G(x_0)$$

x_0 fails to be a local min — exist values smaller than x_0

Since there are values smaller than x_0, it fails to be a local min, instead acts as a global min, meaning it proves our assumption and disproving the 4) statement.

1) Obtaining data samples from the dataset "HW 3 Dataset 1" (which will be sent in an email with that subject line), utilize any or all of the various regression model-fit assessment visual plots and/or graphs, results of statistical hypothesis tests, and numerical regression model accuracy measures, as well as any other applicable diagnostic tools available in R that we have discussed in the course thus far, to help identify and validate a linear regression model that appears to best fit the given data. These will likely include for example direct X-Y variable graphs and residuals vs. fitted value plots implemented in R. You will likely need to try a number of different response and/or predictor variable transformations, for example, using the diagnostic tools to decide which regression models appear consistent with the data and which do not. You should identify one or perhaps two candidate regression models that you believe are most consistent with the given data. Please try to include screenshots of the graphical plots you use as well as quote any relevant R code output. You can or even should include the R code itself as well if you feel it helps to support your argument. Please explain your reasoning as to why the model(s) you propose may be the right one(s) in plain natural language writing.

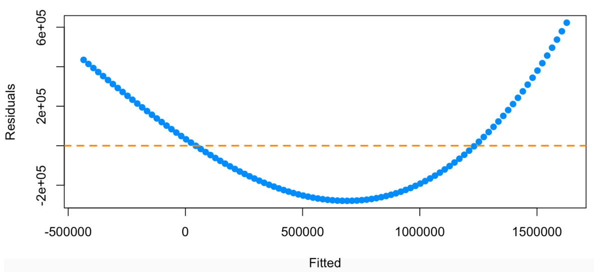
```
dataset1 = read.csv("~/Desktop/PSTAT 126/dataset1.csv")
plot(Y ~ X, data = dataset1,
     xlab = "x",
     ylab = "y",
     pch = 20, cex =2)

dataset1model = lm(Y ~ X, data = dataset1)
summary(dataset1model)
```




```
Call:
lm(formula = Y ~ X, data = dataset1)

Residuals:
    Min      1Q  Median      3Q     Max 
-279166 -224109  -64146  183868  622869 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -538500.6    53216.9  -10.12   <2e-16 ***
X             20820.8      862.9   24.13   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' :

Residual standard error: 249100 on 98 degrees of freedom
Multiple R-squared:  0.8559,    Adjusted R-squared:  0.8545 
F-statistic: 582.2 on 1 and 98 DF,  p-value: < 2.2e-16
```

I plotted dataset1 and found it was a polynomial, the data must be transformed since the fitted vs Residuals plot shows the regression line is not in line with the data. Also, the 1Q and 3Q are unbalanced and the R^2 and adjR^2 are not terrible since they are high but they could be better.

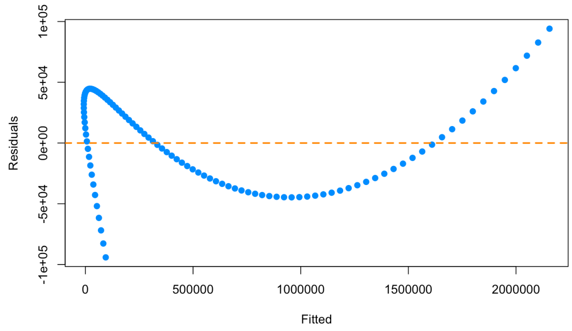
```
plot(fitted(dataset1model), resid(dataset1model), col = "dodgerblue",
     pch = 20, cex = 1.5, xlab = "Fitted", ylab = "Residuals")
abline(h=0, lty = 2, col = "darkorange", lwd = 2)

dataset1_mod = lm(Y ~ X, data = dataset1)
summary(dataset1_mod)
```



```r
#polynomial transform
dataset1_mod1 = lm(Y ~ X + I(X ^ 2), data = dataset1)
summary(dataset1_mod1)
```

```r
plot(fitted(dataset1_mod1), resid(dataset1_mod1), col = "dodgerblue",
     pch = 20, cex = 1.5, xlab = "Fitted", ylab = "Residuals")
abline(h=0, lty = 2, col = "darkorange", lwd = 2)
```

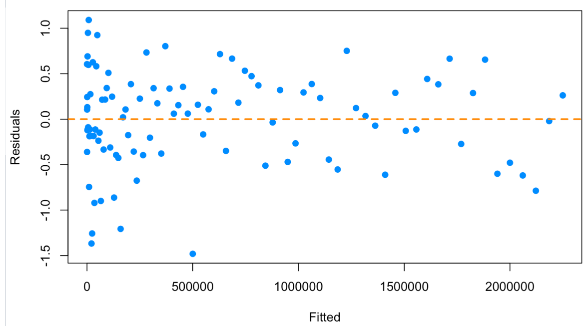
```
Call:
lm(formula = Y ~ X + I(X^2), data = dataset1)

Residuals:
   Min     1Q Median     3Q    Max 
-94109 -33727      1  33728  94110 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 160298.562  13714.413   11.69   <2e-16 ***
X           -14822.207    576.476  -25.71   <2e-16 ***
I(X^2)         327.000      5.146   63.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38350 on 97 degrees of freedom
Multiple R-squared:  0.9966,	Adjusted R-squared:  0.9965 
F-statistic: 1.43e+04 on 2 and 97 DF,  p-value: < 2.2e-16
```



I did a polynomial transformation the quadrants look more balanced and the R^2 and adj^2 are closer to 1 meaning it is a better fit. Although the fitted vs Residuals plot is not great. The variances are not in line with the regression line.

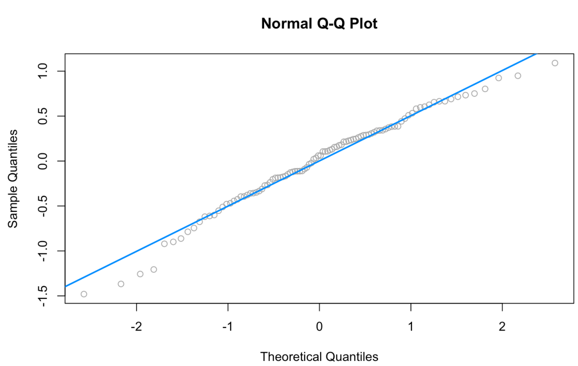
```r
dataset1_mod2 = lm(Y ~ X + I(X ^ 3), data = dataset1)
summary(dataset1_mod2)

plot(fitted(dataset1_mod2), resid(dataset1_mod2), col = "dodgerblue",
     pch = 20, cex = 1.5, xlab = "Fitted", ylab = "Residuals")
abline(h=0, lty = 2, col = "darkorange", lwd = 2)
```

```
Call:
lm(formula = Y ~ X + I(X^3), data = dataset1)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.48039 -0.33762  0.06058  0.34086  1.09026 

Coefficients:
              Estimate Std. Error    t value Pr(>|t|)    
(Intercept)  3.094e+00  1.605e-01  1.928e+01   <2e-16 ***
X           -2.595e-03  4.794e-03 -5.410e-01    0.589    
I(X^3)       2.000e+00  4.260e-07  4.695e+06   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' :

Residual standard error: 0.5252 on 97 degrees of freedom
Multiple R-squared:      1,	Adjusted R-squared:      1 
F-statistic: 7.649e+13 on 2 and 97 DF,  p-value: < 2.2e-16
```



I did another polynomial transformation and this seems like a better fit. The variances in the fitted vs residuals plot are more evenly distributed. The quartiles are balanced meaning it has a Normal distr. and the R^2 and adj R^2 are at 1, suggesting a good fit. The Normal Q-Q plot, plots the sample and theoretical quantiles, since they are in line this suggest Normality of errors.

```r
qqnorm(resid(dataset1_mod2), main = "Normal Q-Q Plot", col = "darkgrey")
qqline(resid(dataset1_mod2), col = "dodgerblue", lwd =2)
```



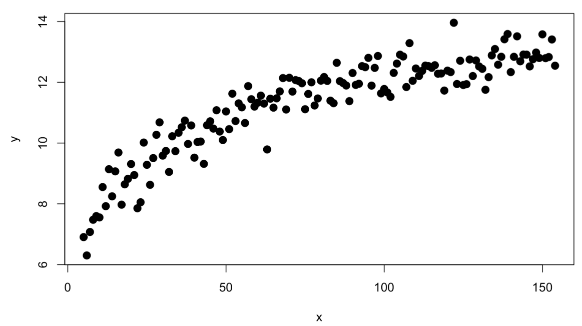
Estimates for parameters:

$\hat{\beta}_0 = 3.094$

$\hat{\beta}_1 - 2.595e^{-03}$

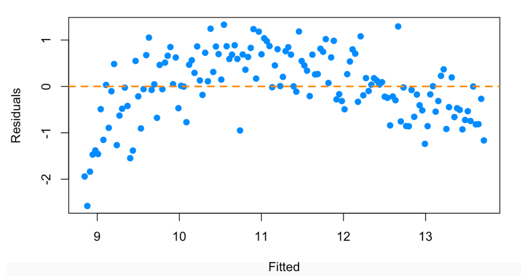
5) Obtaining data samples from the dataset “HW 3 Dataset 2” (which will be sent in an email with that subject line), utilize any or all of the various regression model-fit assessment visual plots and/or graphs, results of statistical hypothesis tests, and numerical regression model accuracy measures, as well as any other applicable diagnostic tools available in R that we have discussed in the course thus far, to help identify and validate a linear regression model that appears to best fit the given data. These will likely include for example direct X-Y variable graphs and residuals vs. fitted value plots implemented in R. You will likely need to try a number of different response and/or predictor variable transformations, for example, using the diagnostic tools to decide which regression models appear consistent with the data and which do not. You should identify one or perhaps two candidate regression models that you believe are most consistent with the given data. Please try to include screenshots of the graphical plots you use as well as quote any relevant R code output. You can or even should include the R code itself as well if you feel it helps to support your argument. Please explain your reasoning as to why the model(s) you propose may be the right one(s) in plain natural language writing.

```
dataset2 = read.csv("~/Desktop/PSTAT 126/dataset2.csv")
plot(Y ~ X, data = dataset2,
     xlab = "x",
     ylab = "y",
     pch = 20, cex =2)

dataset2model = lm(Y ~ X, data = dataset2)
summary(dataset2model)
```



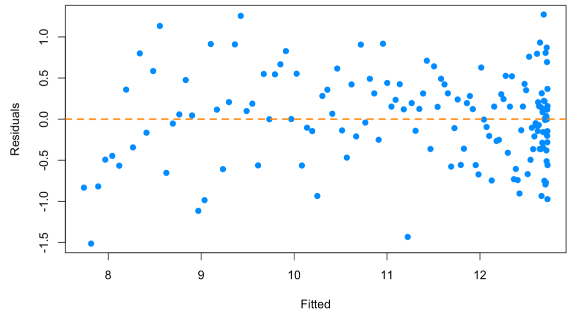

```
plot(fitted(dataset2model), resid(dataset2model), col = "dodgerblue",
     pch = 20, cex = 1.5, xlab = "Fitted", ylab = "Residuals")
abline(h=0, lty = 2, col = "darkorange", lwd = 2)

dataset2_mod = lm(Y ~ X, data = dataset2)
summary(dataset2_mod)
```

```
Call:
lm(formula = Y ~ X, data = dataset2)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.57773 -0.47843  0.02257  0.61561  1.33017 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 8.683026   0.126456   68.66   <2e-16 ***
X           0.032645   0.001397   23.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7408 on 148 degrees of freedom
Multiple R-squared:  0.7868,	Adjusted R-squared:  0.7853 
F-statistic: 546.1 on 1 and 148 DF,  p-value: < 2.2e-16
```

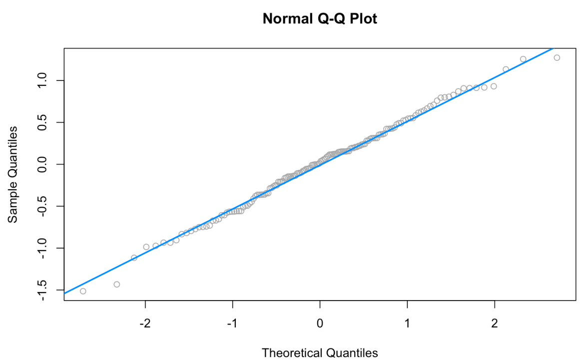


The dataset2 gives an exponential plot. 1Q and 3Q are not balanced, for a normal distribution and the R^2 and adj R^2 could be improved.
While the fitted vs residuals plot does not have a constant variance.

```r
#polynomial transform
dataset2_mod1 = lm(Y ~ X + I(X ^ 2), data = dataset2)
summary(dataset2_mod1)

plot(fitted(dataset2_mod1), resid(dataset2_mod1), col = "dodgerblue",
     pch = 20, cex = 1.5, xlab = "Fitted", ylab = "Residuals")
abline(h=0, lty = 2, col = "darkorange", lwd = 2)

qqnorm(resid(dataset2_mod1), main = "Normal Q-Q Plot", col = "darkgrey")
qqline(resid(dataset2_mod1), col = "dodgerblue", lwd =2)
```




```
Call:
lm(formula = Y ~ X + I(X^2), data = dataset2)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.51432 -0.36277  0.02334  0.34289  1.27257 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.3428633  0.1488919   49.32   <2e-16 ***
X            0.0805797  0.0043039   18.72   <2e-16 ***
I(X^2)      -0.0003015  0.0000263  -11.46   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 

Residual standard error: 0.5401 on 147 degrees of freedom
Multiple R-squared:  0.8874,	Adjusted R-squared:  0.8859 
F-statistic: 579.4 on 2 and 147 DF,  p-value: < 2.2e-16
```



I did a polynomial transformation, the fitted vs residuals has a constant variance and the normal Q-Q plot shows normality of errors. The R^2 and adj R^2 went up closer to 1 now meaning it is a better fit and the 1Q and 3Q are more balanced suggesting a normal distribution. When I did another transformation the graphs didn't change by much and the R^2 and adj R^2 decreased, signaling this one is a better fit.

Estimates

$\hat{\beta}_0 = 7.3428633$

$\hat{\beta}_1 = 0.0805797$