Correlation Assignment:

(i) **Correlation matrix between predictors:**

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{pmatrix} 1.000000e+00 & 9.500000e-01 & -5.730157e-18 \\ 9.500000e-01 & 1.000000e+00 & -1.694362e-16 \\ -5.730157e-18 & -1.694362e-16 & 1.000000e+00 \end{pmatrix} \end{array}$$

Correlation matrix between the predictors and the outcome:

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ y \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & y \end{array} \\ \begin{pmatrix} 1.000000e+00 & 9.500000e-01 & -5.730157e-18 & 8.235151e-01 \\ 9.500000e-01 & 1.000000e+00 & -1.694362e-16 & 7.816903e-01 \\ -5.730157e-18 & -1.694362e-16 & 1.000000e+00 & 1.343833e-01 \\ 8.235151e-01 & 7.816903e-01 & 1.343833e-01 & 1.000000e+00 \end{pmatrix} \end{array}$$

(ii) **Scatter plot between the predictors:**

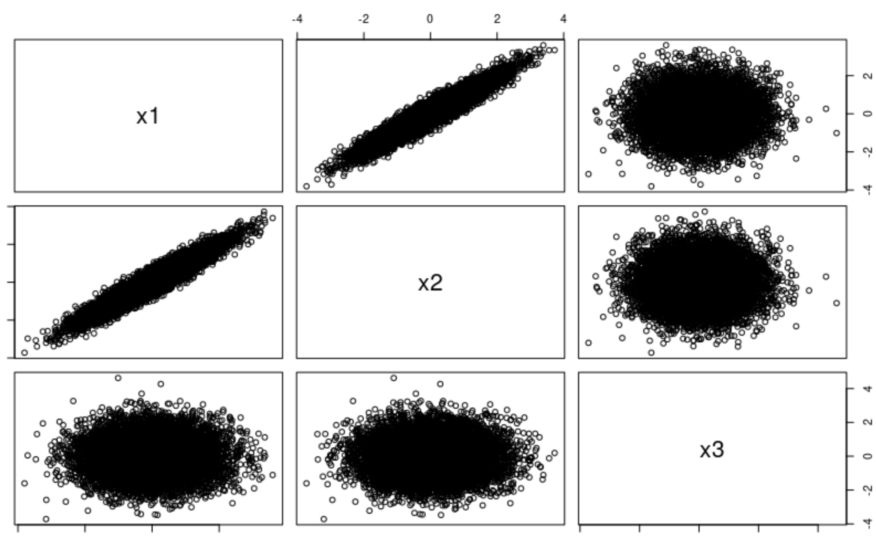

Figure 1: Scatter plot between the predictors.

Interpretation: x1 and x2 have high correlation while they both are not much correlated with x3.

Scatter plot between the predictors and the outcome:

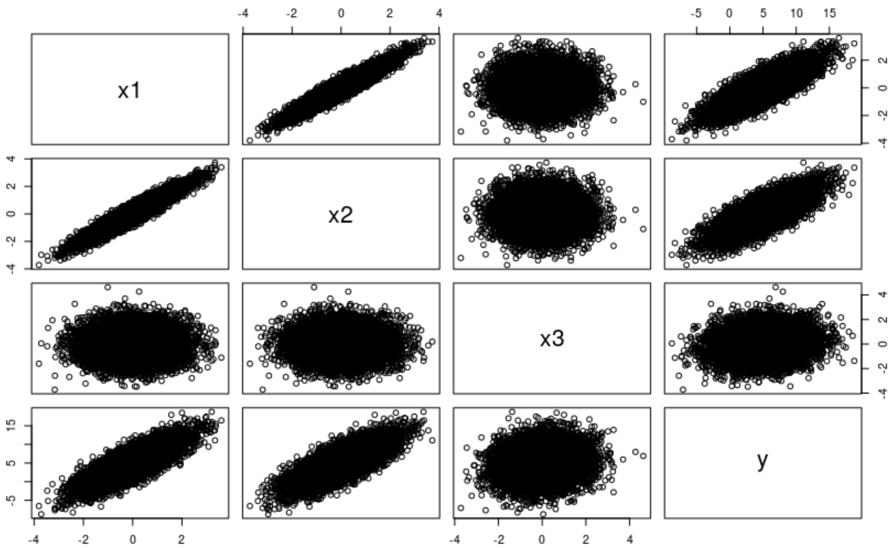

Figure 2: Scatter plot between the predictors and the outcome:

Interpretation: The correlation of x1,x2,x3 with y decreases in the order which they are written here.

(iii) **Based on the correlation values, discuss about the influence of predictors x1, x2, x3 on y:** The higher is the magnitude of the correlation of the predictor, the more influence it will have on the y.

While, for positive correlation between two variables, increase in values of one implies the increase in the value of other and decrease in values of one implies the decrease in the value of other.

On the other hand, for negative correlation between two variables, increase in values of one implies the increase in the value of other and decrease in values of one implies the decrease in the value of other.

Here since x1 has the highers correlation with y which is equal to 0.8.235151, So it is the most influential predictor. Followed by x2 with lesser influence whose correlation with y is 0.7.816903 And x3 has the least influence on y which has correlation with y equals to 0.1.343833.

Also, if two or more variables have same correlation with y we can simply choose one while ignoring others(to reduce the number of predictors) since they will more or less shows similar relationship with y.

(iv) **Fit linear model on the data; Based on the coefficient of the predictors, identify the significant predictors.**

After fitting the linear model on the given data, Q1_data_02, the coefficients obtained are as follows:

```
(Intercept)          x1          x2          x3
 4.99599269  3.03724201 -0.02436174  0.49184846
```

Hence we get the following model:

$$y = 3.03724201 * x1 - 0.02436174 * x2 + 0.49184846 * x3 + 4.99599269$$

The predictor with highest magnitude of coefficient will be most significant as model will be most sensitive to changes of the values of its corresponding variable(predictor).
In our case the magnitude of x1 is highest so it is most important predictor.
Since x2 has very high correlation with x1 and hence its corresponding coefficient has lesser magnitude.
Also, predictor x3 will have lesser effect on the values of y than x1 since its coefficient also has lesser magnitude.

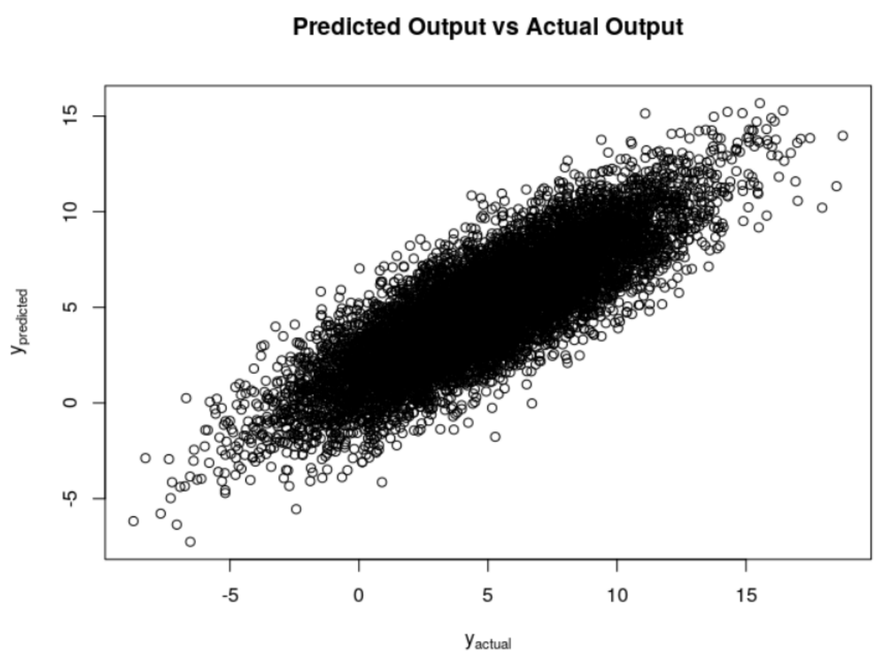

Figure 3: Predicted Output vs Actual Output
Interpretation: The predicted values are quite accurate as the points lie along the direction of line y=x.

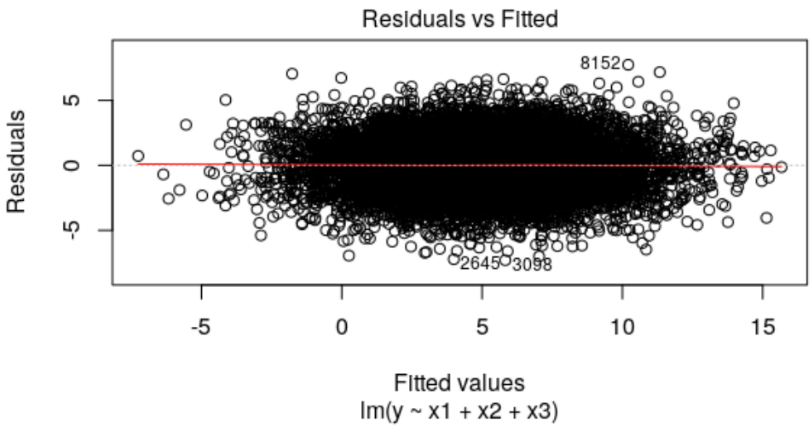

Figure 4: Residual vs Fitted for Linear model

Interpretation: Since the residual spread is equally distributed around horizontal line the relationship between predictors and outcome is not nonlinear.

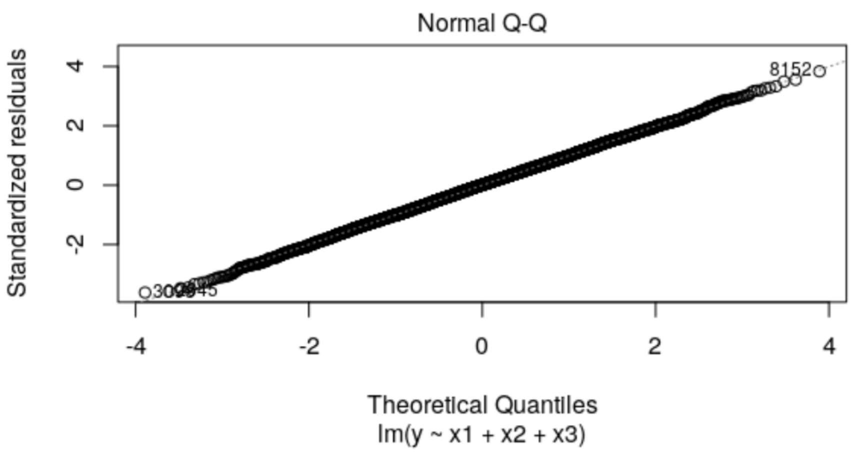

Figure 5: Standardized residuals vs Theoretical Quantifiers

Interpretation: Since the residuals are on straight line it implies that they are normally distributed which indeed is a good sign as they dont deviate much.

Regression - Polynomial Fitting:

$$\text{Q2\_fun\_01: } y = e^{-5(x-0.3)^2} + 0.5\, e^{-100(x-0.5)^2} + 0.5\, e^{-100(x-0.75)^2}$$
$$\text{Q2\_fun\_02: } y = 2 - 3x + 10x^4 - 5x^9 + 6x^{14}$$

(i) **Plot function given in Q2_fun_xx.**

(a)

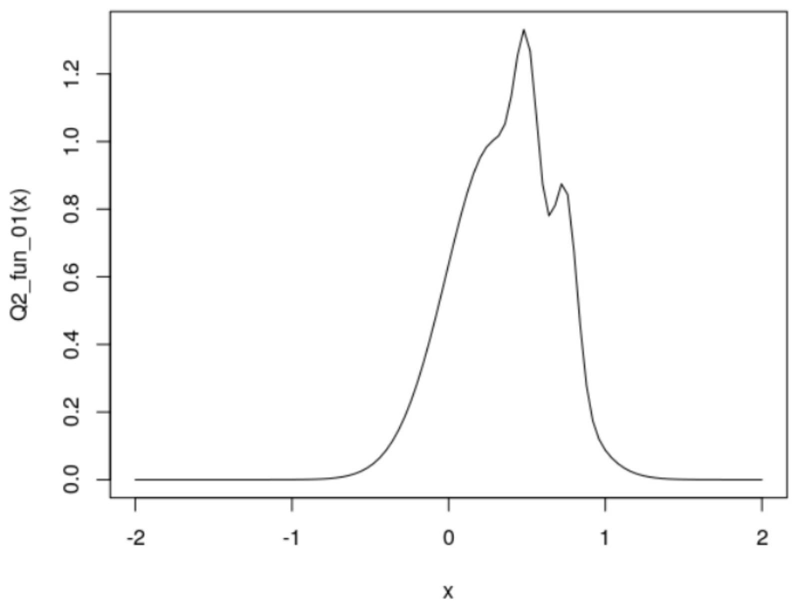

Figure 6: Plot of Q2_fun_01

(b)

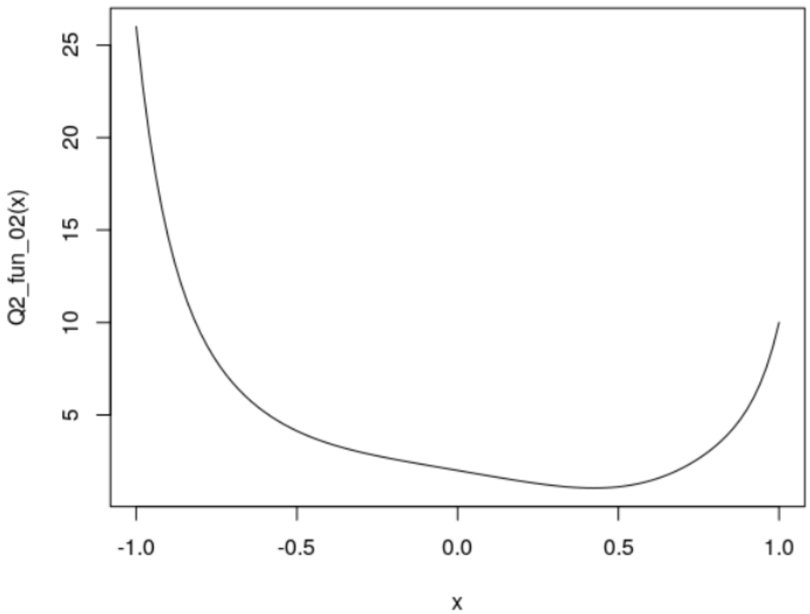

Figure 7: Plot of Q2_fun_02

(ii) **Randomly extract 100 points from the function and add normally distributed noise to the data points to get noisy data $\hat{\mathbf{y}}$.**
For this assignment instead of 100 points, I have sampled 300 points to avoid the error while fitting polynomials of degree 25.
The 300 points were selected randomly from uniform distribution and normally distributed noise from gaussian distribution was added to get the "noisy data".

(iii) Fit polynomial of degrees 8,15,25 to the noisy data.

(a) **For function: Q2_fun_01:**

i. **Coefficients:**
$poly(x\_01, 8)k \implies x^k$
So the coefficients of monomials are given in table with increasing order of degree.

```
Coefficients:
                Estimate
(Intercept)      0.15659
poly(x_01, 8)1 -1.82308
poly(x_01, 8)2 -4.46315
poly(x_01, 8)3 -2.72800
poly(x_01, 8)4  3.80583
poly(x_01, 8)5  2.69265
poly(x_01, 8)6 -1.27333
poly(x_01, 8)7 -2.92933
poly(x_01, 8)8  2.74528
```

Figure 8: Coefficients of polynomial(with degree=8)

```
Coefficients:
                  Estimate
(Intercept)        0.15659
poly(x_01, 15)1  -1.82308
poly(x_01, 15)2  -4.46315
poly(x_01, 15)3  -2.72800
poly(x_01, 15)4   3.80583
poly(x_01, 15)5   2.69265
poly(x_01, 15)6  -1.27333
poly(x_01, 15)7  -2.92933
poly(x_01, 15)8   2.74528
poly(x_01, 15)9  -0.06748
poly(x_01, 15)10  1.51669
poly(x_01, 15)11 -1.46107
poly(x_01, 15)12 -2.15800
poly(x_01, 15)13  0.68931
poly(x_01, 15)14  1.99963
poly(x_01, 15)15 -0.06834
```

Figure 9: Coefficients of polynomial(with degree=15)

```
Coefficients:
                   Estimate
(Intercept)         4.54651
poly(x_02, 25)1   -45.47611
poly(x_02, 25)2    54.67295
poly(x_02, 25)3   -13.71931
poly(x_02, 25)4    23.72181
poly(x_02, 25)5    -6.24675
poly(x_02, 25)6     7.92053
poly(x_02, 25)7    -1.45798
poly(x_02, 25)8     1.62455
poly(x_02, 25)9     1.01939
poly(x_02, 25)10    1.04948
poly(x_02, 25)11    1.75233
poly(x_02, 25)12   -0.74448
poly(x_02, 25)13   -1.23047
poly(x_02, 25)14   -1.08722
poly(x_02, 25)15    1.13913
poly(x_02, 25)16    1.96834
poly(x_02, 25)17    0.70675
poly(x_02, 25)18    0.68519
poly(x_02, 25)19    0.50717
poly(x_02, 25)20    0.35928
poly(x_02, 25)21   -0.06483
poly(x_02, 25)22   -0.82315
poly(x_02, 25)23    1.53192
poly(x_02, 25)24   -0.87526
poly(x_02, 25)25   -1.92978
---
```

Figure 10: Coefficients of polynomial(with degree=25)

ii. **Plots:**

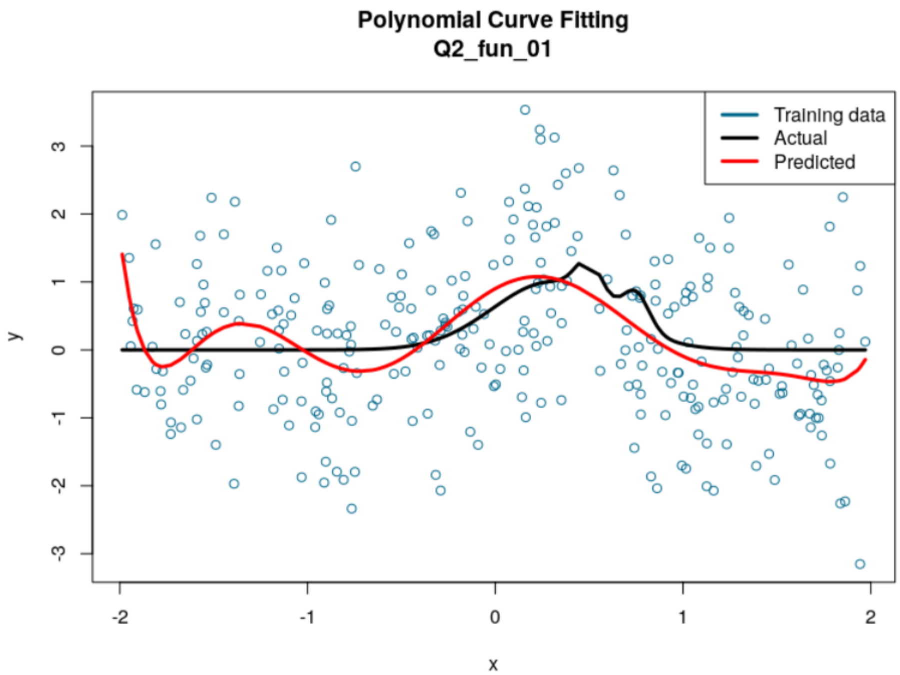

Figure 11: Polynomial curve fitting using degree=8

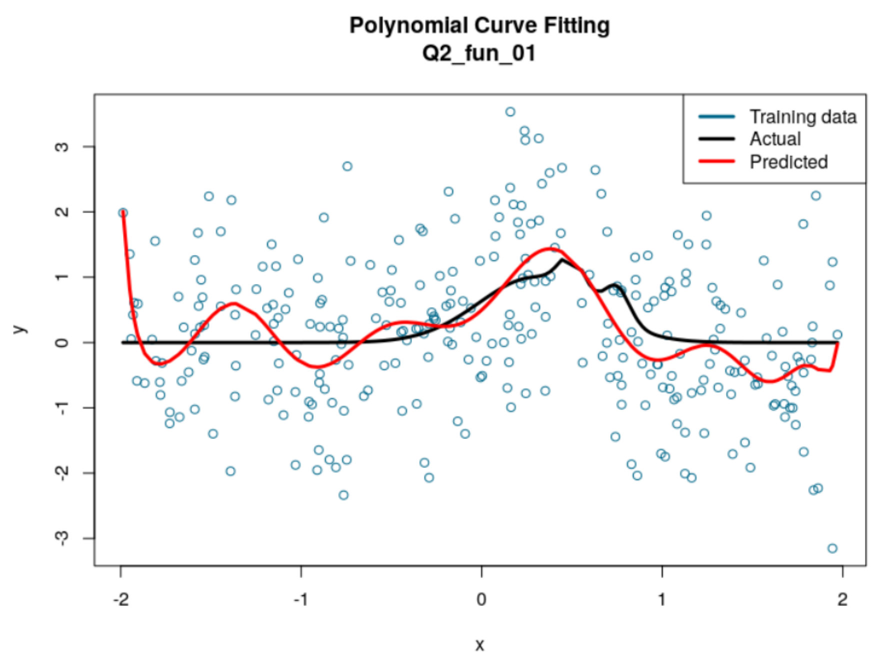

Figure 12: Polynomial curve fitting using degree=15

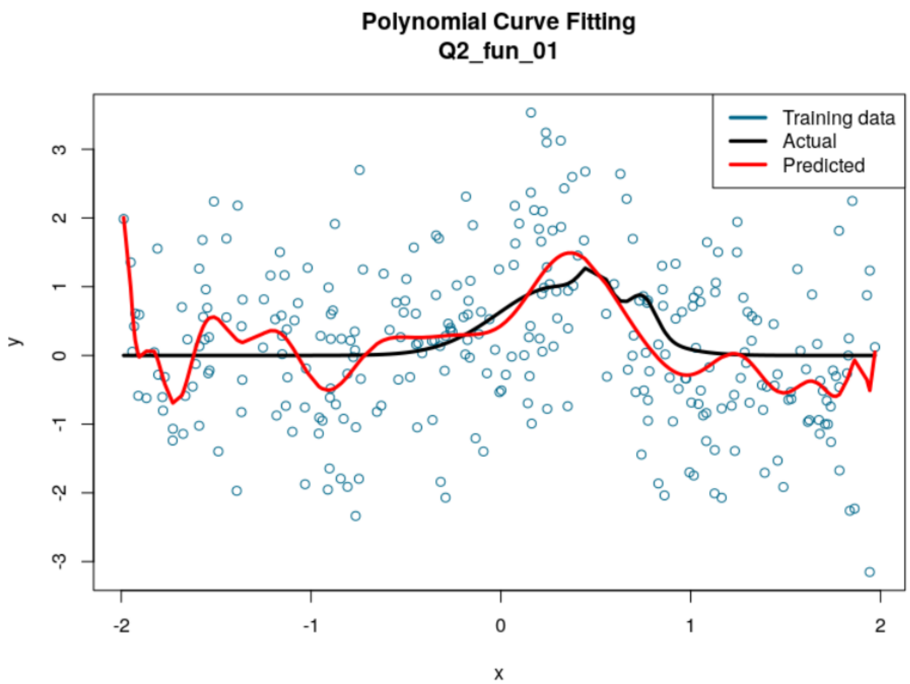

Figure 13: Polynomial curve fitting using degree=25

(b) **For function: Q2_fun_02:**

i. **Coefficients:**
$poly(x\_01, 8)k \implies x^k$
So the coefficients of monomials are given in table with increasing order of degree.

```
Coefficients:
                 Estimate
(Intercept)       4.54651
poly(x_02, 8)1 -45.47611
poly(x_02, 8)2  54.67295
poly(x_02, 8)3 -13.71931
poly(x_02, 8)4  23.72181
poly(x_02, 8)5  -6.24675
poly(x_02, 8)6   7.92053
poly(x_02, 8)7  -1.45798
poly(x_02, 8)8   1.62455
```

Figure 14: Coefficients of polynomial(with degree=8)

```
Coefficients:
                  Estimate
(Intercept)        4.54651
poly(x_02, 15)1  -45.47611
poly(x_02, 15)2   54.67295
poly(x_02, 15)3  -13.71931
poly(x_02, 15)4   23.72181
poly(x_02, 15)5   -6.24675
poly(x_02, 15)6    7.92053
poly(x_02, 15)7   -1.45798
poly(x_02, 15)8    1.62455
poly(x_02, 15)9    1.01939
poly(x_02, 15)10   1.04948
poly(x_02, 15)11   1.75233
poly(x_02, 15)12  -0.74448
poly(x_02, 15)13  -1.23047
poly(x_02, 15)14  -1.08722
poly(x_02, 15)15   1.13913
```

Figure 15: Coefficients of polynomial(with degree=15)

```
Coefficients:
                  Estimate
(Intercept)       0.156589
poly(x_01, 25)1  -1.823081
poly(x_01, 25)2  -4.463149
poly(x_01, 25)3  -2.728002
poly(x_01, 25)4   3.805835
poly(x_01, 25)5   2.692646
poly(x_01, 25)6  -1.273326
poly(x_01, 25)7  -2.929325
poly(x_01, 25)8   2.745281
poly(x_01, 25)9  -0.067479
poly(x_01, 25)10  1.516692
poly(x_01, 25)11 -1.461070
poly(x_01, 25)12 -2.158003
poly(x_01, 25)13  0.689310
poly(x_01, 25)14  1.999630
poly(x_01, 25)15 -0.068342
poly(x_01, 25)16 -0.574287
poly(x_01, 25)17 -0.172896
poly(x_01, 25)18  0.340303
poly(x_01, 25)19 -0.449718
poly(x_01, 25)20  1.123852
poly(x_01, 25)21  0.004221
poly(x_01, 25)22  0.246347
poly(x_01, 25)23  1.427967
poly(x_01, 25)24 -0.813696
poly(x_01, 25)25  0.278017
```

Figure 16: Coefficients of polynomial(with degree=25)

ii. **Plots:**

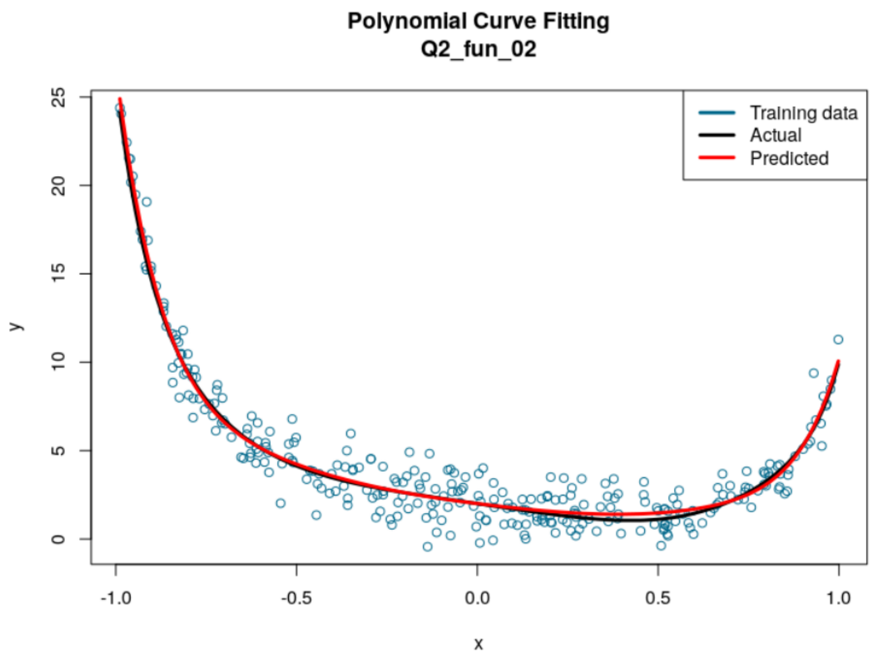

Figure 17: Polynomial curve fitting using degree=8

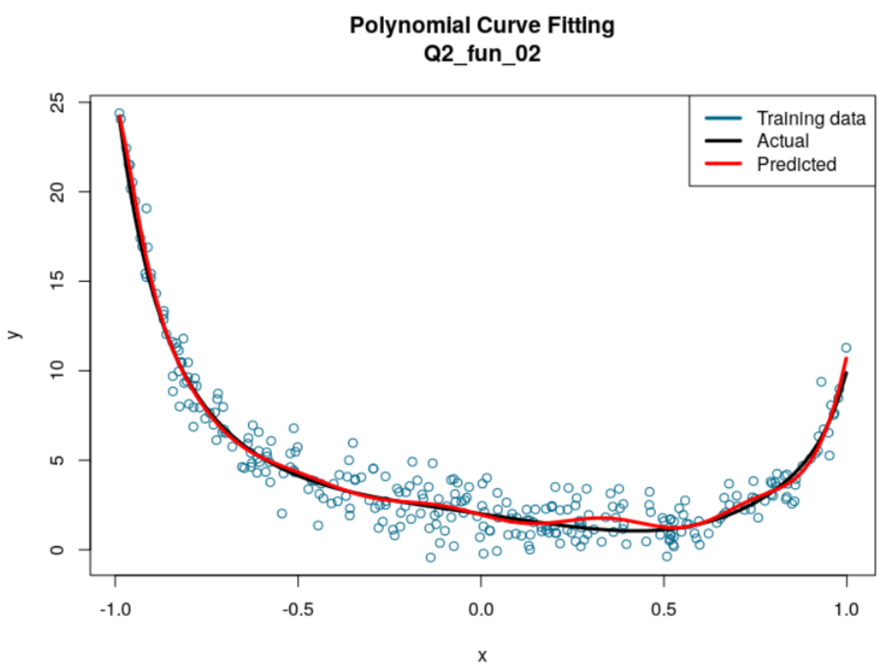

Figure 18: Polynomial curve fitting using degree=15

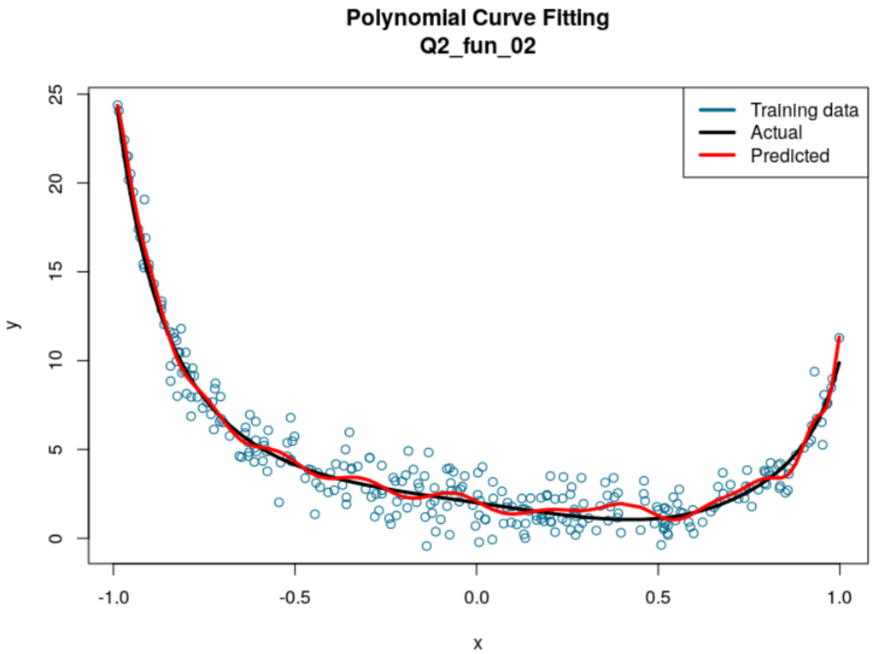

Figure 19: Polynomial curve fitting using degree=25

(iv) **Compute the bias and variance for the models fitted.**
The Bias-Variance is calculated for 10 different models for each degree by keeping same x but varying normally distributed noise to generate 10 different training datasets.

Bias-Variance table for Q2_fun_01:

degree	$Bias^2$	Variance
8	0.012783246	0.02921219
15	0.008590231	0.04354536
25	0.008350728	0.08646242

Bias-Variance table for Q2_fun_02:

degree	$Bias^2$	Variance
8	1.439793e-03	0.0002755726
15	9.713861e-05	0.0004564578
25	9.267311e-05	0.0008999302

(v) **Plot the bias-variance plot.**

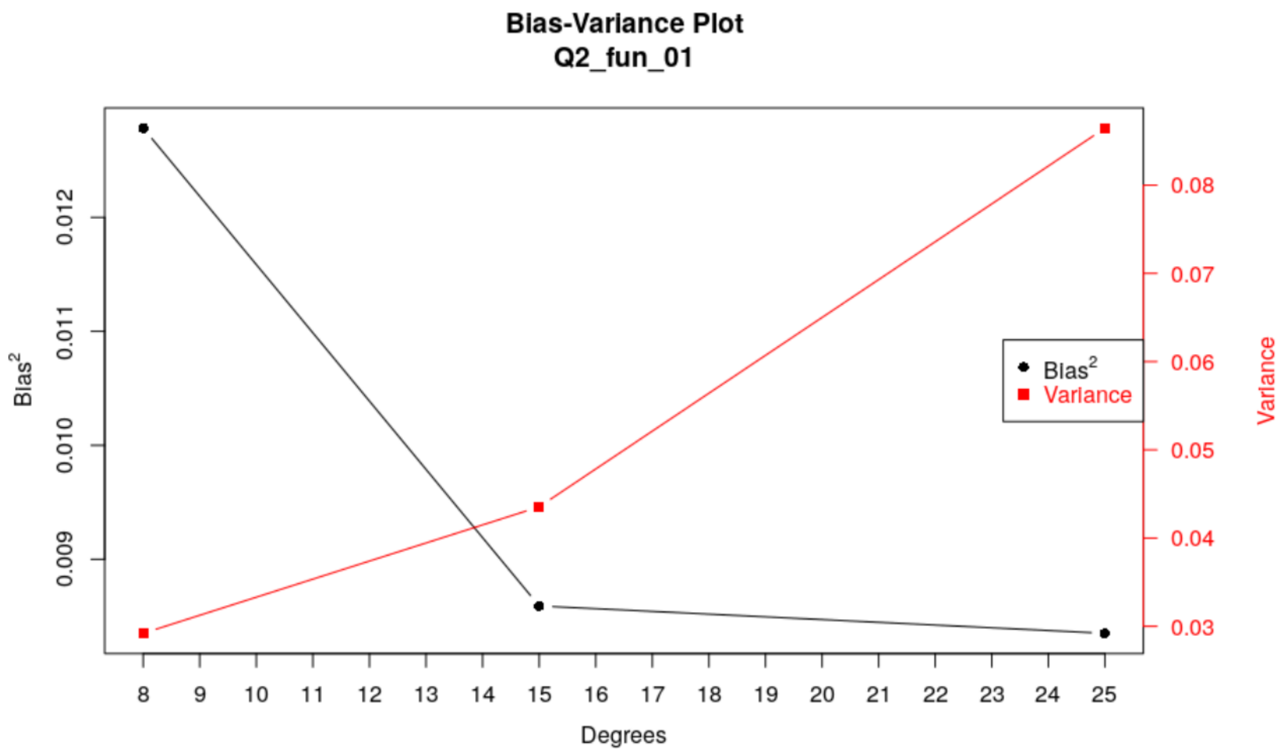

Figure 20: Bias-Variance Plot for Q2_fun_01

Interpretation: With increase in model complexity, the bias is decreasing while error is increasing

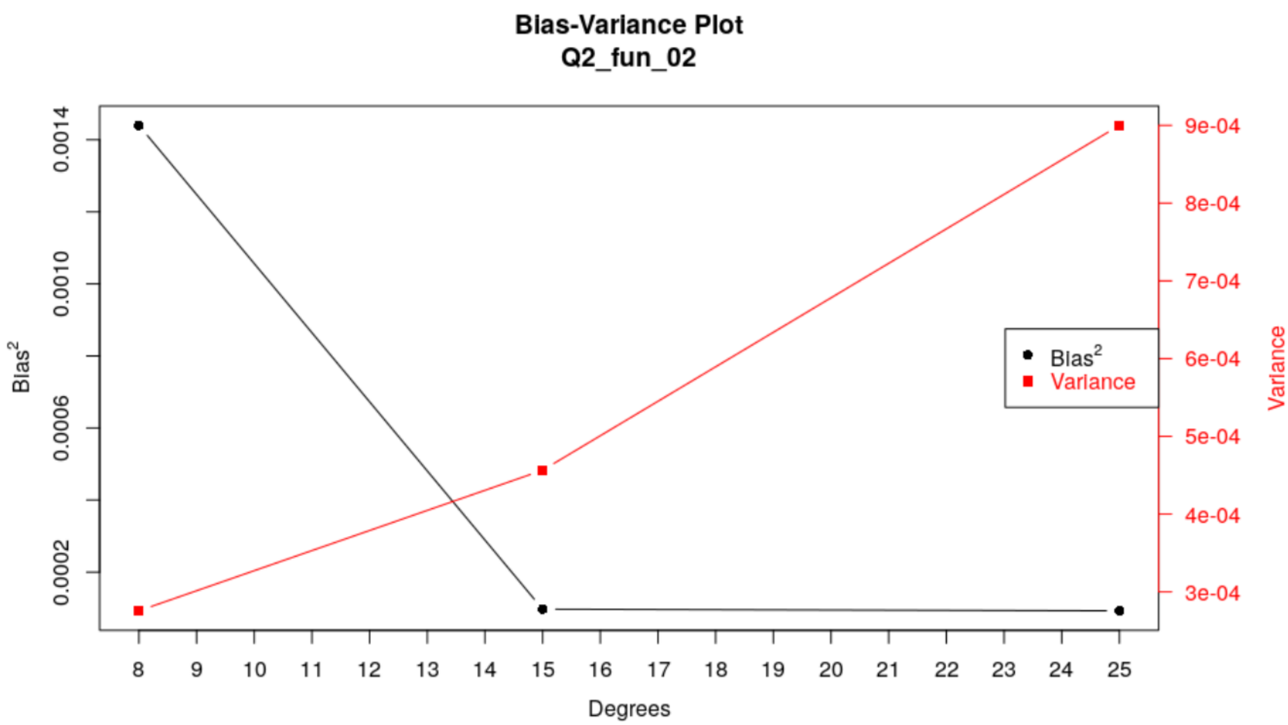

Figure 21: Bias-Variance Plot for Q2_fun_02

Interpretation: With increase in model complexity, the bias is decreasing while error is increasing