Course Project:Non-Linear Regression(MTH-686)

Prerak Agarwal(220818)

November 9, 2024

(a) Computation of LSEs for the three models

Model 1

Our first model is of the form,
$$y(t) = \alpha_0 + \alpha_1 e^{\beta_1 t} + \alpha_2 e^{\beta_2 t}$$
The **Osborne algorithm**, as discussed in the class, is a method typically used in non-linear least squares problems, is applied to optimize parameters, $\alpha_0, \alpha_1, \beta_0, \beta_1$, to minimize the **residual sum of squares (RSS)** between predicted and observed values. The implementation avoids using advanced libraries and relies on basic matrix operations in R.

The optimized parameters, obtained after implementing this algorithm, provides us with these least square estimators(LSEs).

Optimized Parameters	Values
α_0	0.9278005
α_1	0.6910364
α_2	0.6910364
β_1	1.3160677
β_2	1.3160626

Model 2

Our second model is of the form, $y(t) = \frac{\alpha_0+\alpha_1 t}{\beta_0+\beta_1 t}$

The algorithm used in this code is an **iterative optimization** method based on **gradient descent with finite differences** to approximate the gradient. The code approximates the gradient using finite differences since the model is nonlinear:

$$\frac{\partial \text{RSS}}{\partial \beta_j} \approx \frac{\text{RSS}(\beta_j + \epsilon) - \text{RSS}(\beta_j)}{\epsilon}$$

where ϵ is a small value used to perturb each parameter.

Parameters are then iteratively updated using gradient descent:

$$\beta \leftarrow \beta - \text{learning rate} \times \text{gradient}$$

where the learning rate determines the step size for each update in the direction that reduces RSS.

The optimized parameters, obtained after implementing this algorithm, provides us with these least square estimators(LSEs).

Optimized Parameters	Values
α_0	14.437289
α_1	5.724957
β_0	6.308505
β_1	-3.006446

Model 3
Our third model is of the form,
$$y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$$

The **polynomial regression** is a type of regression analysis that models the relationship between an independent variable $y(t)$ and t as an n-degree polynomial. Here, for this model, $n = 4$.
To find the best-fitting curve, we apply the least squares method, minimizing the sum of the squared differences between the observed y values and the predicted values. The coefficients, represented as a column vector, β, are estimated using the normal equation:
$\beta = (X^T X)^{-1} X^T y$
where X is the **design matrix** and y is the **vector of observed values**.
The optimized parameters, obtained after implementing this algorithm, provides us with these least square estimators(LSEs).

Optimized Parameters	**Values**
β_0	2.2860507
β_1	2.0480296
β_2	0.7712057
β_3	0.5524677
β_4	0.4391256

(b) Choice of initial parameters

Model 1
Initial choice of parameters: $\{\alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2\} = \{1, 1, 1, 1, 1\}$
Observing the data, there is a generally increasing trend in y as t increases. This suggests that a model with positive coefficients might be suitable.
Model 2
Initial choice of parameters: $\{\alpha_0, \alpha_1, \beta_0, \beta_1\} = \{10, 2.1, 10, 1\}$
Observing the data, there is a generally increasing trend in y as t increases. Moreover, when t is close to zero, the value of y is 2.3 and when t is close to 1, y is close to 6. Since our model is of form linear/linear, therefore, our initial choice should be such that, at t=0, the ratio of α_0/β_0 should be close to 2, and at t=1, it is close to 6. Thus, this initial is somewhat close to our assumption, and upon testing on different initial parameters, this initial choice provides us with least residual sum of squares.
Model 3
In this model, there is no choice of initial parameters involved.

(c) Finding the 'best fitted model'

Let's compare the values of residual sum of squares for the given three model assumptions.
Model 1: Residual sum of squares=0.1171588
Model 2: Residual sum of squares=0.1149551
Model 3: Residual sum of squares=0.1133745
By comparing the values of residual sum of squares for the given three model assumptions, we conclude that the third model assumption, that is, $y(t) =$

$\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$, provides us with the least residual sum of squares as compared to other two models.
Thus, **model 3 is the 'best fitted model'**.

(d) Estimation of σ^2

The estimate of σ^2 is a crucial statistical parameter that represents the variance of the error terms in our model. It is computed as the variance of the residual terms after we predict the values of Y using the estimated values of the parameters. The estimated value of σ^2 is calculated as follows:

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} (y(t_i) - \hat{y}(t_i))^2$$

Here, **(n=65)** is the **number of data points** and **p** is the **number of parameters** in the model.
Model 1: $p = 5$, thus, the estimate of $\sigma^2 = 0.001952647$
Model 2: $p = 4$, thus, the estimate of $\sigma^2 = 0.001884599$
Model 3: $p = 5$, thus, the estimate of $\sigma^2 = 0.001858599$

(e) Finding the associated confidence intervals based on the Fisher information matrix

1. **Hessian Calculation**: Approximated as $H \approx J^T J$, where J is the Jacobian matrix of the residuals with respect to the parameters.

2. **Regularized Hessian**: The regularized Hessian matrix is computed as

$$H_{\text{reg}} = J^T J + \lambda I$$

   where λ is the regularization parameter and I is the identity matrix.

3. **Fisher Information Matrix**: Calculated as the inverse of the regularized Hessian matrix

$$FIM = H_{\text{reg}}^{-1}.$$

4. **Standard Errors**: Derived as the square root of the diagonal elements of the Fisher Information Matrix.

5. **Confidence Intervals**: Calculated using

$$\hat{\theta} \pm z_{\alpha/2} \times SE(\hat{\theta}),$$

   where $z_{\alpha/2}$ is 1.96 for a 95% confidence interval.

6. **Output**: Confidence intervals provide the lower and upper bounds for each parameter estimate.

Model 1: Confidence Intervals

- α_0: Lower Bound = -3.018742, Upper Bound = 4.874343

- α_1: Lower Bound = -137.909842, Upper Bound = 139.291915
- α_2: Lower Bound = -137.909842, Upper Bound = 139.291915
- β_1: Lower Bound = -137.286206, Upper Bound = 139.918341
- β_2: Lower Bound = -137.286246, Upper Bound = 139.918371

Model 2: Confidence Intervals

- α_0: Lower Bound = -8.373255, Upper Bound = 37.247833
- α_1: Lower Bound = -3.351302, Upper Bound = 14.801215
- β_0: Lower Bound = -3.658321, Upper Bound = 16.275331
- β_1: Lower Bound = -7.758595, Upper Bound = 1.745703

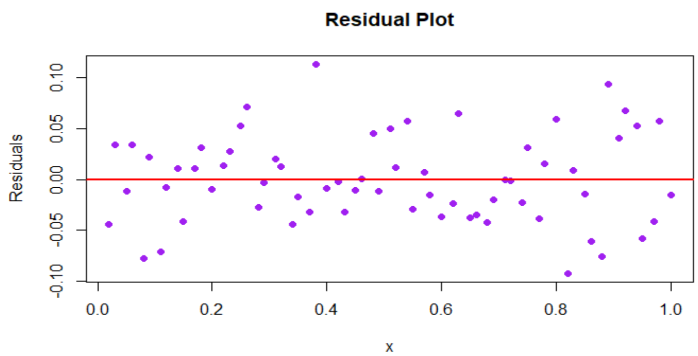
Model 3: Confidence Intervals

- β_0: Lower Bound = 2.226995, Upper Bound = 2.345106
- β_1: Lower Bound = 1.255728, Upper Bound = 2.840331
- β_2: Lower Bound = -2.359504, Upper Bound = 3.901915
- β_3: Lower Bound = -4.056399, Upper Bound = 5.161334
- β_4: Lower Bound = -1.811685, Upper Bound = 2.689936

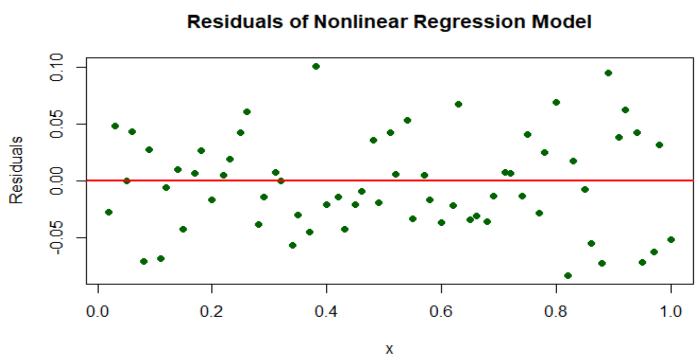
(g) Normality Assumption

In our analysis, we assessed the normality of the residuals derived from our regression model using the **Anderson-Darling normality test**.
A (Test Statistic): Measures the degree of deviation from normality (the lower the better for normality).
p-value: Indicates whether the deviation is statistically significant (a higher p-value suggests the data is normally distributed).
Let's compare the values of A and p for different model assumptions:
Model 1: A = 0.27187, p-value = 0.6606
Model 2: A = 0.29038, p-value = 0.6006
Model 3: A = 0.36945, p-value = 0.4168
By comparing the **Anderson-Darling statistic** to critical values at a significance level of **0.05**, we found no significant deviations from a normal distribution in the residuals of all three models. This outcome supports the key assumption that the residuals closely adhere to a normal distribution, reinforcing the reliability of our regression model and its suitability for subsequent statistical analysis.

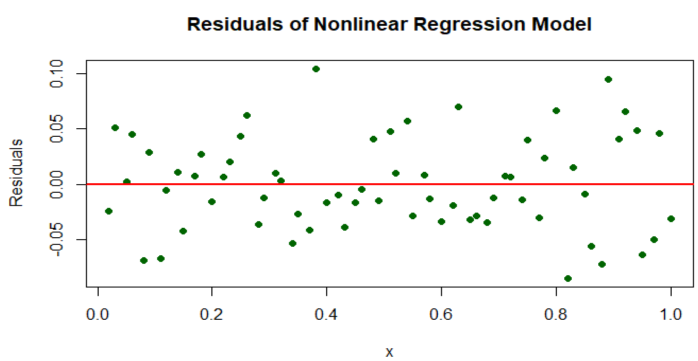
(f) Residual plots and (h) Plotting the observed data points and fitted curve



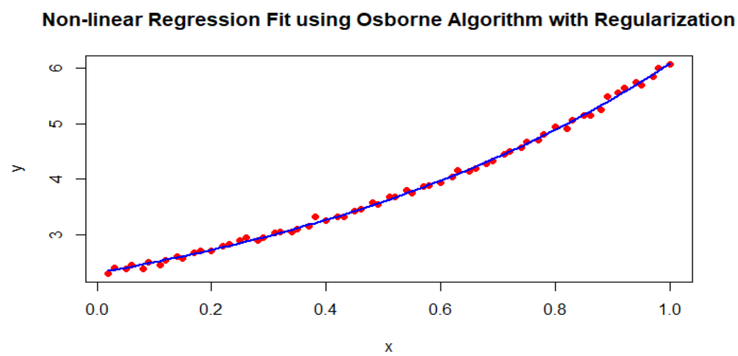
(a) Residual Plot for Model 1



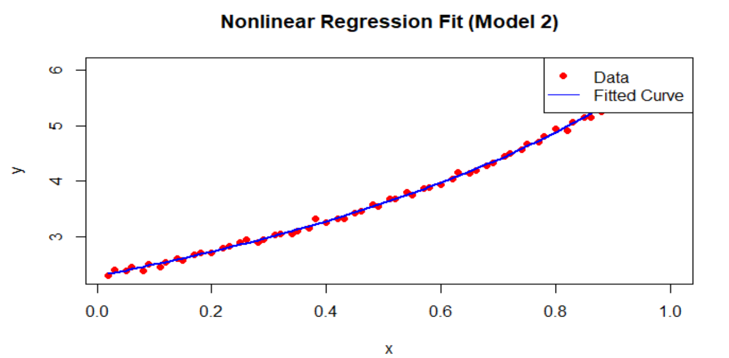
(b) Residual Plot for Model 2



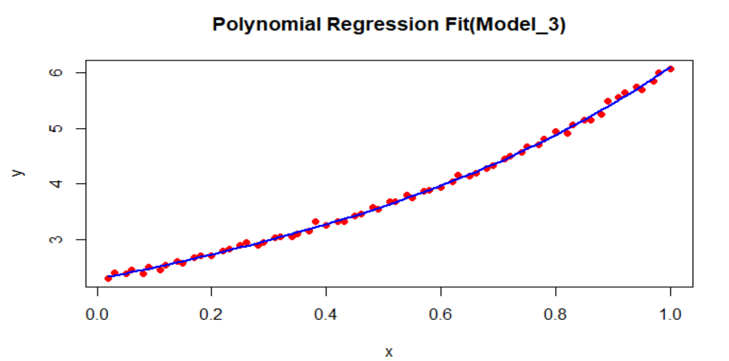
(c) Residual Plot for Model 3



(a) Best Fit Curve Plot for Model 1



(b) Best Fit Curve Plot for Model 2



(c) Best Fit Curve Plot for Model 3