**Team Number: #5**

**Team Captain: Sriharsha Aitharaju**

**Team Members: Daniel Rodriguez**

**Fernando Sosa**

**Kzzy Centeno**

**Robert Law**

**Activity on**

**PART I: (15 Points):**

**Problem 1.1 (5 Points)** The response variable of the observed data and the fitted prediction are listed in the following table.

|  |  |  |
| --- | --- | --- |
| Response (Y) | Model I Prediction ( | Model II Prediction ( |
| 3 | 3.2 | 3.3 |
| 4 | 4.3 | 4.2 |
| 5 | 4.9 | 4.8 |
| 6 | 5.7 | 5.9 |
| 7 | 6.9 | 7.1 |

1. Calculate the sum squared of error of Model I and Model II.

Table

Description automatically generated

1. Calculate the average squared error of Model I and Model II.

Table

Description automatically generated

1. Calculate both and .

Table

Description automatically generated

1. Calculate both and

Table

Description automatically generated

1. Calculate both and

|  |  |  |
| --- | --- | --- |
| Measure | Model I | Model II |
| SSE | **0.06** | **0.0475** |
| ASE | **0.048** | **0.038** |
| R2 | **0.976** | **0.981** |
| MAPE | **1.15%** | **2.15%** |
| MAE | **0.2** | **0.18** |

**Problem 1.2 (10 Points)** Work on Problem 1, Problem 2, and Problem 3 in the Textbook (Chapter 5 on Page 219)

Text

Description automatically generated

Text, letter

Description automatically generated

Text

Description automatically generated

Text

Description automatically generatedGraphical user interface

Description automatically generated

Text

Description automatically generated with medium confidence

1. Using *n* for observations and *k* for observations randomly split up in the data set we use the equation *n* / *k*. We can now use the average of *k* to obtain the MSE estimate. The observations we used to split up is used as a validation set.
2. i.
   * K-Fold Validation can be applied to almost all learning algorithms
   * The computation time for K-Fold Validation is approximately K times of the computation time of “Validation Set Approach” that is acceptable for most learning algorithms
   * 5-Fold or 10-fold validation have been shown to yield the test error rate that suffer neither from excessively high bias nor from very high variance.
   * Disadvantages are that there may be overestimation when based on a few observations

ii. The LOOCV

* The LOOCV method can have a shorter computation time but in special cases

**PART II Programming (15 Points)**

**Data Used:** “House\_Prices\_PRED.CSV” with three variables: ID, House\_Price (observed value), and P\_House\_Price (Model Predicted Value).

**Problem 2.1 (0 Points)** Read the CSV file “House\_Prices\_PRED.CSV”

**Problem 2.2 (3 Points)** Write a program to calculate the sum squared of error and the average squared error of the Model (i.e., P\_House\_Price).

**Problem 2.3 (3 Points)** Write a program to calculate the R2 of the Model (i.e., P\_House\_Price).

**Problem 2.4 (3 Points)** Write a program to calculate the MAPE of the Model (i.e., P\_House\_Price).

**Problem 2.5 (3 Points)** Write a program to calculate the MAE of the Model (i.e., P\_House\_Price).

**Problem 2.6 (3 Points)** Write a program to produce a residual plot with residual on the Y-axis and observed value (House\_Price) and to impose a loess line on the graph.