Department of Mechanical Engineering

ME 781: Engineering Data Mining and Applications

Assignment-1

With reference to the given data set: a1-data-set.csv

1. What are your general observations about this data set?
2. Decide your strategy for imputing the missing values.
3. Create a new data set with missing values appropriately substituted
4. Based on your analysis, decide what kind of model you will fit over this data if the goal is to use the given data to predict 'y' given 'x'
5. Fit the model and note its parameters.
6. Does the model accurately represent the data? Justify your answer.
7. Calculate the following values for the predicted 'y' values w.r.t the given 'y' values:
    - RMSE, MAE (Mean of absolute error)
8. Plot y-given v/s y-predicted: what are your observations?
9. Plot the residuals (y-predicted – y-given) v/s 'x': what are your observations? Based on this, what do you have to say about your model?

Note:

- Submit your answers to the assignment submission point in Moodle.
- This is one of a set of 5 Problems that you will have to solve. Submission of all assignments will fetch you max 5 marks.
- The Test scheduled on Nov-8-2017 will assume you have done this assignment.