

Assignment 1: Linear Models
Generalized Linear Models (GLM)
Course Number: MA687
Semester 2, 2019-20, IIT Guwahati

Total Marks: 40

Instructions:

- Read the questions carefully before answering them.
- Marks for each question are given on the right hand side.
- You may do your assignment in any statistical software. However, I prefer R. **You need to submit the codes (in email) along with a report (hard copy) preferably in pdf.**
- All the codes need **detailed** documentation. I will randomly choose some codes and will run in my computer to check. If the code is written in other software than R, then I will call you to run that.
- **Deadline to submit the hard copy of the report (codes should be sent in email) of this assignment: In the class, 5PM 24 February 2020.** You may also submit it before the deadline.
- You will get zero if you miss the deadline.

In the email, you will get a data on ‘Concrete Compressive Strength’ with the file name “Concrete.Data” along with a data information file “Concrete.Readme”. Read the data information file carefully to understand the data. The column “ID” is not an input variable, you should not consider this in your model.

1. Fit a multiple linear regression to the above data considering all the input variables. The information of the output variable is given in the information file. [5]
2. Explain the results obtained from the statistical software along with the diagnostic results/plots. I expect your explanation to be detailed along with interpretations. [10]
3. Based on your diagnostics plots’ interpretations, do you want to recommend any changes in the model? If so do the changes along with the reasons and fit the model again. If no change required then support your arguments. [5]

4. In the same email, you will get another data as “add_columns”. Merge this new data with the previous data. Now fit a multiple linear regression on the merged data with all the input variables. Explain your results. Comment on the newly added variables and whether you want to keep them in the model or not. Justify your answer in either case. [10]
5. You have decided that you will only allow three input variables in the model to make it simple. Choose the most appropriate three input variables and justify your answer along with results. [10]