

TAMS 65 Assignment 7

Multiple linear regression - All subsets regression

This assignment deals with multiple linear regression as well as All subsets regression.

Instruction

- You are recommended to use software MATLAB to answer questions.
- Attach the project in **pdf file** and name it as **Project.pdf**.
- Make a [detailed report](#) on **2 assignments**.
- Give [only solutions](#) to the rest of **5 assignments**. Note: I have marked this part of questions in **blue**. That is, you only need to show solutions to the blue part of questions for the rest of 5 assignments.
- **Submit first version** of your report to your **teaching assistants** [not later than May 1, 2020](#).
- **Submit final version** of your report to the lisam: **Lisam - Submissions**.
- **Deadline** for submissions is at [23:00 May 15, 2020](#). Note: The submission entrance will [open](#) at [0:00 May 7, 2020](#).
- **All codes** that you will need are given either in Lectures or 7 assignments.

All subsets regression

People want to analyze the chemical yield for an industrial process, but they are not sure which explanatory variables they should use in regression model. In a feasibility study with 30 observations, it is desirable to study which variables should be included in the model. Let

$$\begin{array}{ll} Y = & \text{measure of chemical yield,} \\ x_1 = & \text{amount of catalyst,} \\ x_2 = & \begin{cases} 1, & \text{preprocessing 2;} \\ 0, & \text{otherwise} \end{cases} \\ x_3 = & \begin{cases} 1, & \text{preprocessing 3;} \\ 0, & \text{otherwise} \end{cases} \\ x_4 = & \text{humidity i \%}, \\ x_5 = & \text{percentage of oxygen in the surrounding environment,} \\ x_6 = & \text{time in seconds for the process,} \\ x_7 = & \text{square of time of process,} \\ x_8 = & \text{temperature i } ^\circ C. \end{array}$$

Download the files **all_subset.m**, **Assignment7.m** and **Assignment7A.m**. Input the codes in the **Command Window** or **Editor** window.

Questions

- (a) Open the file **Assignment7.m**, then **run** it. Perform a regression analysis with all 8 variables. Calculate the coefficient of determination R^2 and do a residual analysis. Note: Here you need to define response variable y and variables $x_i, i = 1, \dots, 8$.
- (b) Propose a model by applying all subsets regression. Calculate the coefficient of determination R^2 and do a residual analysis. You may need to borrow the code from the file **Assignment7A.m** or you can just run the file **Assignment7A.m**.
- (c) Compare your proposed model in (b) with the full model using all 8 explanatory variables. Is the full model significantly better?

Instructions to All subsets regression

First we consider models with all possible combinations of all possible explanatory variables. For example, we need to consider 2^k models if we have k explanatory variables.

Then we compare these models in various ways. For example, we can choose the model with minimum MSE or maximum adjusted R^2 .

You may use the following code which is used for the model with minimum MSE.

```
model= all_subset(y,X);  
model
```