Assignment 2 10 Marks 26th March 2018

You have to answer either Q1 or Q2.

Q1. For the following datasets given in the Table, you have to report the total squared error from leave-one-out cross validation. (https://www.cs.cmu.edu/~schneide/tut5/node42.html)

You have to submit the following filled table and your code in google classroom.

Dataset #	Dataset Name	Total Squared Error
1	Airfoil Self-Noise	
2	Combined Cycle Power Plant	
3	Challenger USA Space Shuttle O-Ring	
4	Computer Hardware	
5	ISTANBUL STOCK EXCHANGE	
6	NoisyOffice	
7	Parking Birmingham	
8	Physicochemical Properties of Protein Tertiary Structure	
9	Relative location of CT slices on axial axis	
10	Residential Building Data Set	
11	SML2010	
12	Stock portfolio performance	
13	Tamilnadu Electricity Board Hourly Readings	
14	Wine Quality	
15	Yacht Hydrodynamics	

The datasets are available at https://archive.ics.uci.edu/ml/datasets.html (Filter the datasets by 'Regression') on the left top panel.

Please be aware that institute plagiarism rules apply.

Assignment 2 10 Marks 26th March 2018

Q2. This is a continuation of the previous assignment. You used H-index as input and Impact Factor as output in the previous assignment.

But there are other input factors like SJR, Total docs, Total ref.s etc. In this assignment, you have to consider more input factors to improve the results from the previous assignment. Note that, taking all possible factors may not give you the best results. You have to choose the optimal set of inputs (by trial and error) that yields the best results.

This is an open ended research problem. You are free to show the results in any fashion you choose. But you have to report the following bare minimum …

Combo Number	Mean Absolute Error	Mean Squared Error
Combo #1		
Combo #2		
.		
.		

In each Combo, you have mention the combination of inputs you have considered (order of inputs do not matter in regression). Try to achieve the best results.

You have submit the results and script files.