# **Machine Learning 310 Problem set 1**

Spring 2015

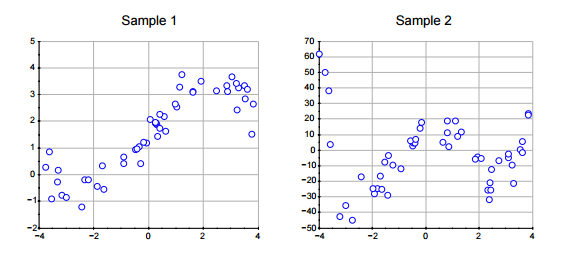
Instructor: Dipankar Ray

## Data files

The questions make use of four data files:

|  |  |
| --- | --- |
| **Train** | **Test** |
| hw1\_sample1\_train.txt | hw1\_sample1\_test.txt |
| hw1\_sample2\_train.txt | hw1\_sample2\_test.txt |

Each data file contains n = 50 lines. Each line contains exactly 2 numbers representing the X and Y coordinates of points generated by adding noise to a secret function. The files hw1 sample2 train.txt and hw1 sample2 train.txt are only useful for the last question. For verification, here is a graphical representation of both datasets:



**Question 1**:

Implement linear least square curve fitting, on both of the two samples provided. Provide plots showing the points and the fitted curve for both datasets using the following bases.

* Linear regression:
* Cubic regression:
* Cubic splines:

The notation is a shorthand for .

**Question 2**:

Download the two testing sets provided on the homework page and report the testing set MSE achieved by the models trained above.