

## Assignment 2

### Assignments on Incomplete and Noisy data:

1. Use Linear regression and multiple regression to compute the **missing values** of an attribute.
2. Smooth the values of an attribute to **remove the noise** by binning methods [consider equidepth bins]: (a) bin means (b) bin medians (c) bin boundaries.

### Assignment on Discretization:

3. Discretize a continuous valued attribute of a dataset into K-discrete values using following procedures:
  - (a) Histogram Analysis (Equal width & Equal depth)
  - (b) Cluster Analysis (use K-Means clustering algorithm and the clustering algorithm developed in Assignment 1)
  - (c) 3-4-5 rule based natural segmentation
  - (d) Entropy-Based Discretization