# Hand-on Exercise 1

## Guidelines

(1) Each group consists of at most two students. Members in the same team will receive the same grade. (Please give your group member names to TA.)

(2) The deadline for turning in your report is February 19 (Thursday) 1:00 PM.

## Assignment

(1) Each team's task is to unveil the embedded information of a given data set.

(2) The data set is posted on the course page.

(3) Each team may use MATLAB/Python (or others) as the development platform.

(4) The performance will be judged by the quality of solutions obtained, the associated analysis and discussions.

(5) Major works include:

(i) Data Clustering (30 pts): Use the k-means algorithm and apply both Elbow rule method and Silhouette method to conduct your investigation.

(ii) Data Regression (30 pts): For each cluster, find the hidden linear relationship of the data points in the cluster.

(iii) Display of Results (20 pts): Identify the centroid of each cluster and the range of the cluster (the longest distance from the centroid to a data point in the cluster); draw the linear relationship (you found for each cluster) passing the centroid over the whole range.

(iv) Analysis and Discussion (20 pts): Analyze your findings and discuss related issues.

## Report Format

(1) Title and Team Members

(2) Background and Objectives

(3) Data Clustering Work and Results

(4) Data Regression Work and Results

(5) Graphic Display of the Results

(6) Analysis and Discussions

(7) Attachment of Computer Program/Codes, Inputs and Outputs