

Hand-on Exercise Description

1. Rules

- (1) Each group consists of at most two students. (Please give the names to TA.)
- (2) Members in the same team will receive the same grade.
- (3) The deadline for turning in your report is October 31 (Thursday) 3:00PM.

2. 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 choose MATLAB or Python as the development platform.
- (4) The performance will be judged by the quality of solutions obtained and the analysis of results.
- (5) Major works include (i) Data Clustering (5 pts) - use the "k-mean algorithm" and "elbow rule method" to determine the total number of clusters and data points in each cluster; (ii) Data regression (5 pts) - for each cluster, find the hidden linear relationship represented by data points in the cluster; (iii) Display of Results (3 pts) - identify the centroid of each cluster and the range of 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 (2 pts) - analyze your findings and discuss related issues.

3. Report: Suggested structure:

- (1) Report Title and Authors
- (2) Background and Objective of the report
- (3) Data clustering work and results
- (4) Data regression work and results
- (5) Graphic display of the results obtained
- (6) Analysis and Discussion
- (7) Attachment of the computer program/codes