



CEG426/626

DATA CLUSTERING AND ANALYSIS

No. Units:

4

Prerequisites:

CS400/600 & MTH253

Textbook:

Data Clustering: Theory, Algorithms, and Applications

By G. Gan, C. Ma, and J. Wu

2007 ASA-SIAM (American Statistical Association and the Society for Individual and Applied mathematics)

Contents:

1. Data types
2. Scale conversion
3. Data standardization and transformation
4. Data visualization
5. Similarity and dissimilarity measures
6. Hierarchical clustering
7. Center-based clustering
8. Search-based clustering
9. Graph-based clustering
10. Clustering evaluation
11. Applications of clustering

Purpose of course:

This course will introduce to students the basic tools for analysis and visualization of exploratory data.

Learning goals:

Students will learn the basic concepts in data analysis and clustering and practice some of the concepts through computer implementation.

Projects and exams:

There will be four 30-minute quizzes and four programming assignments. A quiz will cover materials taught in past two weeks in class, and a programming assignment will require implementation of one of the concepts learned in class. A programming assignment is anticipated to take from 15 to 20 hours of work.

Grading policy:

Quizzes will worth 50% and programming assignments will worth 50% of the overall grade. Grades will be curved based on overall class performance. However, the following grades are guaranteed at the minimum:

A: [91..100], B: [81..90], C: [71..80], D: [61..70], E: [0..60].

Calendar:

Programming Assignment 1	Assigned: Due:
Quiz 1	On:
Programming Assignment 2	Assigned: Due:
Quiz 2	On:
Programming Assignment 3	Assigned: Due:
Quiz 3	On:
Programming Assignment 4	Assigned: Due:
Quiz 4	On:

