Cluster analysis (undergrad or grad)

Textbook:


Tentative class schedule:

1) Introduction: what is a cluster; visualization techniques (chapters 1 & 2)

2) Measure Distance (Chapter 3)
Distance measure for categorical data; Distance measure for continuous data; inter-group proximity measure; weighting

3) Hierarchical clustering: (Chapter 4).
Divisive vs. agglomerative approaches; The dendogram; Choice of the number of cluster; Large data set approach;

4) Optimization techniques: (Chapter 5)
dissimilarity matrices, Cluster criteria; K-means and k-median algorithms

5) choosing the number of groups & evaluating clusters (assigned readings)
Silhouette width, Calinski-Harabasz index
Kaufman & Rousseuw, Finding Groups in Data: An Introduction to Cluster Analysis, chapter 2.

6) Finite mixture models (Chapter 6)
Maximum likelihood estimation; Bayesian analysis of mixture;

7) Model-Based cluster analysis (Chapter 7)