Tutorial 1 Unsupervised Learning- Clustering

# Lecture

## Cluster Analysis

In this week’s lecture, we explored the concepts about cluster analysis and finish off with a ‘by-the-hand’ example on how to do a cluster analysis. Of course in reality, we will always use a statistical program to do a cluster analysis and produce a dendrogram, but it is important that you know how to do a small example by hand. This way you will understand the mathematics behind the algorithms.

**Textbook Reading**

10 Unsupervised Learning 373  
10.3 Clustering Methods . . . . . . . . . . . . . . . . . . . . . . . 385  
10.3.1 K-Means Clustering . . . . . . . . . . . . . . . . . . 386  
10.3.2 Hierarchical Clustering . . . . . . . . . . . . . . . . . 390  
10.3.3 Practical Issues in Clustering . . . . . . . . . . . . . 399

Note: there are online lectures and powerpoints that follow these sections of the textbook.

In a group of two, discuss the following topics

* What is the difference between supervised and unsupervised learning?
* Give an example of where unsupervised learning can be used to support decision-making.
* What is the objective of cluster analysis?
* List two clustering methods covered in this week’s lecture.
* What is dissimilarity?
* What is K-means clustering?
  + How does it work?
* How does K-means clustering differ from K-medoids clustering?
* List the pros and cons of K-means and K-medoids clustering.
* What is hierarchical clustering?
* What is a dendrogram?
* Explain the underlying algorithm of hierarchical clustering with single linkage.
  + How does it differ from hierarchical clustering with complete linkage?
* Give an example of when correlation is a better choice for measuring similarity compared to Euclidean distance.

# Tutorial

The tutorial for this week will focus on Cluster Analysis. It is unlikely you will be able to finish all the questions listed on the task sheet during the tutorial time. Please ensure you attempt at least one question for each topic.

## Independendent Learning

### Labs

10.5 Lab 2: Clustering . . . . . . . . . . . . . . . . . . . . . . . . 404  
10.5.1 K-Means Clustering . . . . . . . . . . . . . . . . . . 404  
10.5.2 Hierarchical Clustering . . . . . . . . . . . . . . . . . 406

R code for these labs is available from the ISLR website. There are also youtube clips which follow the labs fairly closely, but not always exactly.

### Exercise 8, Chapter 10

### Exercise 9, Chapter 10

## Peer-to-Peer Problems

### Exercise 2, Chapter 10

### Exericse 3, Chapter 10