First day

Title: Introduction to R and statistical analysis

Abstract:

On the first day of the lecture, we explain the basic usage of R and the statistical terminology that will be necessary in the subsequent lectures. We also present selected techniques from descriptive statistics to summarize and visualize data.

Second day

Title: Linear Regression

Abstract:

On the second day, we learn how to construct a statistical model to explain or predict a quantitative variable by other variables based on data. We focus on Linear Regression Analysis, which is one of the most standard and important tools in statistical analysis.

The lecture on the second day will basically follow the contents in Chapter 3 of "An Introduction to Statistical Learning" (Authors: G. James, D. Witten, T. Hastie, R. Tibshirani, Publisher: Springer, 2013).

Third day:

Title: Classification

Abstract:

On the third day, we consider data with classification labels. The aim of analysis is to construct a statistical model to predict correct classification labels based on other variables contained in the data. We present two statistical methods, Logistic Regression and Linear Discriminant Analysis (LDA), to accomplish this purpose.

The lecture on the third day will basically follow the contents in Chapter 4 of "An Introduction to Statistical Learning" (Authors: G. James, D. Witten, T. Hastie, R. Tibshirani, Publisher: Springer, 2013).

Fourth day:

Title: Unsupervised Learning

Abstract:

On the fourth day, we learn some statistical methods to “compress” data when the data contain so many variables that it is difficult to summarize or visualize them by standard techniques. In particular, we explain the method called Principal Component Analysis (PCA), which aims at constructing a few linear combinations of the variables to explain whole of the data as efficient as possible. If we have time, we will also present the method called Cluster Analysis, which aims at constructing a reasonable classification of the variables induced by the data.

The lecture on the fourth day will basically follow the contents in Chapter 10 of “An Introduction to Statistical Learning” (Authors: G. James, D. Witten, T. Hastie, R. Tibshirani, Publisher: Springer, 2013).