

Multivariate Statistical Analysis

2020 Spring

School of Industrial Management Engineering, Korea University

1. Course Description

- This module aims to provide students with the theoretical and practical knowledge and skills to obtain, modify, and analyze a large amount of data from various sources.
- Topics covered in this module include supervised learning algorithms (classification & regression) such as artificial neural networks, unsupervised learning algorithms (clustering) such as K-Means clustering and hierarchical clustering, and anomaly detection algorithms such as local outlier factor (LOF) and Isolation Forest.
- This module comprises lectures and lab exercises with R to develop the practical skills.
- Students are strongly required to take Data Mining (IMEN321) before taking this course.
- Students must have knowledge about R language (prerequisite: Programming Language for Data Analysis, 2nd semester for the first grade)

2. Lecturer

- Pilsung Kang, Innovation Hall 801A, 02-3290-3383, pilsung_kang@korea.ac.kr
- Course homepage: <https://github.com/pilsung-kang/multivariate-data-analysis>

3. Time, Place, and Textbook

- Time: Tue 14: 00~15:15 (75 min.), Thu 14:00~15:15 (75 min.)
- Place:
 1. New Engineering Hall B-113
- Textbook: No single textbook is required. Lecture slides and associated materials (R script, data sets, etc.) will be provided through the blackboard and course homepage.

4. Lecture Module

- 2~3 lecture units consists one module: 1~2 unit: classroom lecture(s), 1 unit: R exercise

5. Introduce Yourself

- ✓ Submit your self-introduction slide (max. 5 pages) to the lecturer via E-mail by the end of the 2nd week.
- ✓ Required information: Name, department, e-mail, cell phone number, recent photo(s)

6. Assessments

- 2 exams (midterm exam 20%, final exam 30%, total 50%): 2 pages of cheating papers are allowed
- 4 assignments (40%)
- Attendance (10%)

7. Schedule

Week	Topics
Mar. 03/05	Orientation
Mar. 10	Logistic Regression: Classroom lecture 1
Mar. 12	Variable Selection 1: Classroom lecture 2
Mar. 17	Variable Selection 2: Classroom lecture 3
Mar. 19	R Exercise for Logistic Regression and Variable Selection
Mar. 24	Decision Tree: Classroom lecture 1
Mar. 26	Ensemble 1 (Backgrounds): Classroom lecture 2
Mar. 31	Ensemble 2 (AdaBoost): Classroom lecture 3
Apr. 2	Ensemble 3 (Random Forest, GBM): Classroom lecture 4
Apr. 7	R Exercise for Decision Tree and Ensemble
Apr. 9	Clustering 1 (KMC, HC): Classroom lecture 1
Apr. 14	Clustering 2 (DBSCAN): Classroom lecture 2
Apr. 16	R Exercise for Clustering
Apr. 21	No Class
Apr. 23	Midterm Exam
Apr. 28	Artificial Neural Network 1 (Perceptron): Classroom lecture 1
May. 07	Artificial Neural Network 2 (MLP): Classroom lecture 2
May. 12	Artificial Neural Network 3 (Bagging): Classroom lecture 3
May. 14	R Exercise for Clustering
May. 19	No Class (IE Conference)
May. 21	Anomaly Detection 1 (Density-based): Classroom lecture 1
May. 26	Anomaly Detection 2 (LOF): Classroom lecture 2
May. 28	Anomaly Detection 3 (IForest): Classroom lecture 3
Jun. 02	R Exercise for Anomaly Detection
Jun. 06	No Class
Jun. 09	Final Exam