Multivariate Statistical Analysis
2021 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
   - Youtube video lecture playlist
     1. https://youtube.com/playlist?list=PLetSlH8YjIfWKLpMp-r6enJvnk6L93wz2

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 (Offline)
     2. Zoom (Online)
   - 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
   - Online video lecture
     1. Students are required to watch the assigned videos before the class

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
   - 1 final exam (34%): Open book
   - 8 assignments (7% each, 56% total)
   - Q&A Participation (10%)

7. Schedule

Date	Topics
3/2 (Tue)	Orientation
3/4 (Thu)	Introduction to Multivariate Data Analysis Part 1
3/9 (The)	Introduction to Multivariate Data Analysis Part 2
3/11 (Thu)	Multiple Linear Regression
3/16 (Tue)	Multiple Linear Regression: R Exercise
3/18 (Thu)	No class
3/23 (Tue)	Logistic Regression: Formulation, Learning, Interpretation
3/25 (Thu)	No class
3/30 (Tue)	Classification Performance Evaluation
4/1 (Thu)	Logistic Regression: R Exercise
4/6 (Tue)	Dimensionality Reduction: Overview, FS, BE, SS
4/8 (Thu)	Dimensionality Reduction: GA, Shrinkage Methods
4/13 (Tue)	Dimensionality Reduction: R Exercise
4/15 (Thu)	Decision Tree
4/20 (Tue)	Decision Tree: R Exercise
4/22 (Thu)	Artificial Neural Network: Perceptron, Gradient Descent
4/27 (Tue)	Artificial Neural Network: MLP
4/29 (Thu)	Ensemble Learning: Overview, Bagging
5/4 (Tue)	Ensemble Learning: Random Forests
5/6 (Thu)	Ensemble Learning: AdaBoost
5/11 (Tue)	Ensemble Learning: Gradient Boosting Machine
5/13 (Thu)	Ensemble Learning: R Exercise
5/18 (Tue)	Association Rule Mining
5/20 (Thu)	Association Rule Mining: R Exercise
5/25 (Tue)	Clustering: Overview, K-Means Clustering
5/27 (Thu)	Clustering: Hierarchical Clustering, DBSCAN
6/1 (Tue)	Clustering: R Exercise
6/3 (Thu)	No class
6/8 (Tue)	QA Session for Final Exam
6/10 (Thu)	Final Exam