Business Analytics
2020 Fall
School of Industrial Management Engineering, Korea University

1. Course Description

- This course provides advanced topics in statistical/machine learning fields. This course also aims at fostering graduate students to have Python programming skills to implement the introduced algorithms during the lecture.

2. Topics

- Dimensionality reduction: forward/backward/stepwise selection, genetic algorithm, principal component analysis (PCA), multidimensional scaling (MDS), locally linear embedding, ISOMAP, t-SNE, etc.
- Kernel-based learning: support vector machine (SVM), support vector regression (SVR), Kernel Fisher discriminant analysis (KFDA), Kernel principal component analysis (KPCA), etc.
- Novelty detection: Gaussian density estimation, mixture of Gaussians, Parzen window density estimation, 1-SVM, SVDD, local outlier factor (LOF), iForests, etc.
- Ensemble learning: Bagging, AdaBoost, Gradient Boosting Machine (GBM), XGBoost, CatBoost, Random Forests, etc.
- Semi-supervised learning: Self-training, Generative models, semi-supervised SVM, Graph-based SSL, Co-training, etc.

3. Prerequisites

- Linear algebra: Eigenvector/Eigenvalue, Singular value decomposition (SVD), Lagrangian multiplier
- Basic machine learning algorithms: Linear regression, Logistic regression, Decision tree, etc.
- Optimization: Gradient descent, expectation-maximization (EM) algorithm, etc.

4. Time, Place, Lecturer, and Course Homepage

- Time: Tue/Thu 15:30 ~ 17:00
- Place: 218, New Engineering Hall
- Lecturer: Pilsung Kang, Innovation Hall 801A, 02-3290-3383, pilsung_kang@korea.ac.kr
- Course homepage
  1. Github: https://github.com/pilsung-kang/Business-Analytics-IME654-
  2. Youtube: https://www.youtube.com/watch?v=ytRmxBvyGG0&list=PLetSlH8YjIfWMdw9AuLR5ybkVvGcoG2EW
  3. Slack Channel: ime654-koreauniv.slack.com

5. Textbook

- Necessary materials such as journal/conference articles or book chapters will be provided.

6. 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)

7. Assessments

- 1 final exam (25%)
  1. Three pages of cheating sheets are allowed
- Youtube videos (25%)
  1. Students must upload a short video (max 5 minutes) that reviews the topic covered in the scheduled lecture within 48 hours after the class.
  2. A student must explain what he/she learns in the class to his/her partner.
- 5 paper reproduction posts (50%, 10% each)
  1. Students are required to reproduce one excellent (at least good) paper with Python for each topic category (5 reproductions in total).
  2. The contents of the selected paper and corresponding python code should be explained in a single web page post.

8. Schedule

Week	Topics
1	Orientation
2	Dimensionality reduction: forward/backward/stepwise selection, genetic algorithm, PCA
3	Dimensionality reduction: MDS, ISOMAP, LLE, t-SNE
4	Kernel-based learning: SVM
5	Kernel-based learning: SVR, KFDA, KPCA
6	Novelty detection: Gauss, MoG, Parzen
7	Novelty detection: k-NN, LOF, 1-SVM, SVDD
8	No Class
9	Novelty detection: PCA-based, Clustering-based, iForest, Robust random cut forsts
10	Ensemble learning: Bagging, Random Forests
11	Ensemble learning: AdaBoost, GBM
12	Ensemble learning: XGBoost, CatBoost
13	Semi-supervised learning: Self-training, Generative models
14	Semi-supervised learning: SS-SVM, Graph-based SSL
15	Semi-supervised learning: Co-Training, (Re)MixMatch, FixMatch
16	Final Exam