Final Project

Objectives

- Apply what you have learned to analyze real dataset
- Interpret the results of data analysis
- Present your work

Application Datasets

- **UC Irvine Machine Learning Repository**
  https://archive.ics.uci.edu/ml/datasets.php

1. Arrhythmia Dataset (279 attributes, 452 instances)
http://archive.ics.uci.edu/ml/datasets/Arrhythmia

Distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups.

2. Steel Plates Faults Dataset (27 attributes, 1941 instances)
https://archive.ics.uci.edu/ml/datasets/steel+plates+faults

Steel plates faults are classified into 7 types.

3. PM2.5 Dataset (86 attributes, 52854 instances)
https://archive.ics.uci.edu/ml/datasets/PM2.5+Data+of+Five+Chinese+Cities

Hourly PM2.5 data in five Chinese cities.

Application Datasets

4. Dermatology Dataset (33 attributes, 366 instances)
https://archive.ics.uci.edu/ml/datasets/Dermatology

Erythemato-squamous diseases are classified into 6 types.

5. Teaching Assistant Evaluation Dataset (5 attributes, 151 instances)
https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation

TA's performance scores are divided into three categories.

6. Concrete Compressive Strength Dataset (9 attributes, 1030 instances)
https://archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength

Concrete compressive strength and related variables.

Tasks

- Choose one from the given datasets
- Formulate an appropriate problem
- Propose a strategy/procedure to solve the problem
- Perform the analysis (training & test)
- Write a report

Report

- 1. Power Point slides (pdf format)
  - Title page: Course name, project title, your name, student ID, date
  - Organization: Background, problem, strategy/methods, justification, results, conclusions, discussions/what you have learned
  - Page limit: 12 slides (including title page); extra slides will be ignored.
  - DON'T include codes in the slides.

- 2. Appendix (pdf format): codes used in your data analysis

Submission deadline: November 28, Sunday @10:00 PM

Requirements

- This is an individual assignment.
- Discussion with other students is allowed.
- Reading references related with the chosen dataset is allowed.
- BUT you are required to do independent data analysis and report writing.
- You can ask TA for help ONLY if you have problem in downloading the dataset and importing the data to R.