IT541 – Introduction to Machine Learning

Course Objective

Data Science is the study of extracting information from data. Addressing the needs of the industry requires skills in data processing and data analytics spanning a wide area of subjects ranging from Statistics to Machine Learning. This course will introduce some of the basic concepts, techniques and tools that are required to solve problems widely seen in data analytics with Machine Learning tools. The course wil start with a short review on inferencial Statistics and exploratory data analysis. The focus in the study of these subjects will be breadth, rather than depth, and practical *examples* will be used in applications of regression and classification techniques to a wide variety of problems in predictive data analytics.

Projected Outline

- **Introduction and Terminology**
- **Python Recap & Data Preprocessing / Exploratory Analysis with Pandas**
- **Linear Regression**
- **Scikit Learn – A Deeper Look**
- **Model Accuracy and Bias-Variance Trade-off**
- **Introduction to Classification: Logistic Regression**
- **Regularization and Model Tuning**
- **K-Nearest Neighbors**
- **Tree Based Methods and Bootstrapping**
- **Support Vector Machines (Optional)**
- **Natural Language Processing (Optional)**
- **An unsupervised approach: Dimensionality Reduction and PCA (Optional)**

Grading

- Labs : 40%
- Course Project : 30%
- Final Exam : 30%

Course Project

A problem will be submitted for solution. In the final week of the course you'll present your solution. Projects are required to be in groups of at least 2.
Requirements: Notebook and Presentation files + Short Presentation in class

Presentation Details (10 mins at most)

- *Problem Description*
- *Exploratory Analysis (25%)*
- *Data Preprocessing (25%)*
- *Model Tuning (25%)*
- *Results (25%)*
- *Conclusion*

Supplementary Books

- **Introduction to Statistical Learning**
  - James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.
  - Full Book Page:
    - https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf
  - Scikit-Learn Website

Schedule

Class Hours:
Saturdays 13:00 pm, Online
Thursdays 19:00 pm, Online

Contact

Altuğ Tanaltay
Office : FENS – G001B
Office Tel : 0(216) 4839561
Email : atanaltay@sabanciuniv.edu