DA 514
Machine Learning I

Instructor : H. Sait Ölmez
Office : FENS 1088, Tuzla Campus
Contact : Email: olmez@sabanciuniv.edu / Tel: (216) 483 9578

Teaching assistant : Yalçın Can Kılıç, yalcinkilic@sabanciuniv.edu

Course schedule

Term : Spring Term, 2021-2022 academic year
Start/End : January 22nd, 2022 – March 7th, 2022
When : Monday 19:00 – 22:00 / Saturday: 13:00 – 16:00

Course description and outline

This course will introduce the fundamental concepts, techniques and tools that are required to solve problems widely seen in Machine Learning. The course will start with a short review of data pre-processing and exploratory data analysis. The focus of the course will be on both breadth and depth, and practical examples will be used in applications of classification techniques to a wide variety of problems in predictive analytics.

Students should gain a reasonably solid foundation in supervised learning algorithms, model building and model evaluation. They will understand the theory behind these methods, and be able to apply them correctly and appropriately to analyze the data, build a model, and interpret the results.

Outline (tentative)

- Introduction to predictive analytics
- Exploratory data analysis and visualization
- Pre-processing data
    - Data cleansing, data preparation, treatment of missing data, variable transformation, feature selection, feature engineering
- Bias-variance tradeoff and success metrics
    - Concepts of under- and over-fitting, bias and variance in a Machine Learning model, how to measure the success of a model, model selection techniques, model tuning
- Supervised Learning algorithms
    - Logistic Regression
    - Linear Discriminant Analysis (LDA)
    - Naïve-Bayes
    - k Nearest Neighbors
    - Decision Trees
    - Ensemble Learning
        - Bagging algorithms: Bagged Trees, Random Forest, Extra Trees
        - Boosting algorithms: Adaboost, Gradient Boosting, and XGBoost
    - Support Vector Machines (SVM)
        - Linear and kernel-based SVM
    - Introduction to Artificial Neural Networks (if time permits)

Course Format

Lectures will be supported with Machine Learning models and programming work using Python. You

are expected to work with Jupyter Notebooks (or Colab) during the course. All submissions (assignments, midterm, take-home etc.) are electronic and must be in the .ipynb format uploaded to SUCourse+ as Jupyter Notebooks.

Course Notes
Course lectures will be uploaded to SUCourse+ in the pdf format. Code snippets and other Python programs we use in class will be available on SUCourse+.

Textbook
There is not an assigned textbook for this class. There is a vast amount of text and video material on the Internet. I will, however, suggest the following books as reference:

- An Introduction to Statistical Learning: with Applications in R, corrected 7th printing, 2017, by G. James, D. Witten, T. Hastie and R. Tibshirani, http://www-bcf.usc.edu/~gareth/ISL/
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd edition, 2019, by Aurelien Geron
- An Introduction to Machine Learning with Python: A Guide for Data Scientists, 4th edition, 2018, by Andreas C. Müller and Sarah Guido
- Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and Tensorflow2. 3rd edition, by Sebastian Raschka and Vahid Mirjalili

Assessment
Course grading will be based on the total number of points that the participant receives on quizzes, exams and homework assignments. Tentative breakdown of the overall course grading is:

- 1 Midterm exam 35%
- 1 Final exam (take-home exam) 40%
- 3 Homework assignments (2 HW + 1 HW-Project) 25%

Software:
We will use Python and its relevant libraries (scipy, stasmodels, numpy, scikit-learn) throughout the course (exercises, homework assignments and take-home exams).

Course Communication
All course-related announcements will be made through SUCourse. Lectures, handouts, solution sets and Python Notebooks will be posted on SUCourse. E-mail is my preferred way of communication if you need to reach me. I check my mail on a regular basis and reply relatively fast.

Collaboration
You are welcome and even encouraged to discuss the problems in your homework assignments with your fellow classmates but are expected to submit your own work prepared in your own words. You are also expected to know, understand, and comply with the ethical standards of the university and stay away from any sort of dishonesty (cheating, copying somebody else's work, etc.).

Disclaimer
The information provided in this syllabus is tentative and is subject to change. Any modifications to the syllabus will be announced via email or SUCourse, or you'll be verbally notified in class.