Syllabus (Spring 2021-2022)
DA517 Machine Learning II

Meeting Times

Saturday	13:00 - 16:00
Wednesday	19:00 - 22:00

Instructor

İnanç Arın e-mail: inanc.arin@sabanciuniv.edu

Teaching Assistant

Yalçın Can Kılıç email: yalcinkilic@sabanciuniv.edu

Course Description, Aim and Content

In the scope of this course we will cover various machine learning and data mining methods, namely clustering and classification applied to different data types. We will also look at recommendation systems and personalization where data mining models play an important role. In this scope, we will have a deep dive into the semantic relations between instances, and these semantic relation techniques will be used for text mining as well. Students are expected to understand the fundamental theory behind each technique, as well as implementing them using an environment such as Python. Some machine learning concepts will be recovered by applying them in Python. This course also aims to improve practical programming skills of the students for implementing basic/advanced machine learning techniques. Therefore, we will be developing lots of Python applications to reinforce our theoretical knowledge.

Course Content[1]

1. Gradient Descent Techniques
    a. Batch Gradient Descent
    b. Stochastic Gradient Descent
    c. Mini-Batch Gradient Descent
2. Some Machine Learning Concepts (Recap with Python)
    a. End to end machine learning application, Imbalanced dataset
    b. Overfitting/Underfitting, Learning Curves
    c. Handling missing values
3. Clustering Algorithms
    a. Partition Based Clustering Algorithms
    b. Hierarchical Clustering Algorithms
    c. Density Based Clustering Algorithms and Outlier Detection with DBSCAN
        i. Special Nearest Neighbor Techniques to improve DBSCAN efficiency
    d. Evaluation of Clustering Algorithms
4. Dimensionality Reduction
    a. Linear Projection
        i. PCA (Incremental & Kernel PCA)
        ii. SVD
    b. Manifold Learning
        i. t-SNE
5. Data Mining for Recommendation Systems and Personalization
    a. Content Based Recommendation Systems
    b. Collaborative Filtering
    c. Hybrid Approaches
    d. Latent Factor based Techniques for Recommendation Systems (SVD)
6. Text Mining (Introduction to NLP)
    a. Converting unstructured data into structured data
    b. Stop words
    c. Stemming, Lemmatization
    d. Applying SVD to Text Mining
7. Artificial Neural Network with Keras (MLP ➡ DNN)
    a. Sequential API
    b. Functional API

---

[1] Instructor may change the contents and the order of the topics during the semester

      c.   SubClassAPI (maybe)

      d.   Wide & Deep Neural Networks

      e.   Callbacks

8.   Advanced end to end machine learning application development (if time permits)

Grading

- Take-Home Exams   60% (Group and/or individual exercises)
- Final Exam   40% (Both theoretical and practical with Python)

References

- Rajaraman, A., & Ullman, J. D. (2011). Mining of massive datasets. Cambridge University Press.
- Patel, A. A. (2019). Hands-On Unsupervised Learning Using Python: How to Build Applied Machine Learning Solutions from Unlabeled Data. O'Reilly Media.
- Hapke, H. M., Lane, H., & Howard, C. (2019). Natural language processing in action.
- Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2019.
- Data Mining Concepts and Techniques, Jiawei Han & Micheline Kamber & Jian Pei, Morgan Kaufmann (3rd edition)
- Scikit-Learn: Machine Learning in Python https://scikit-learn.org/stable/index.html
- Some other additional resources will be shared in our course Google Drive folder.

Announcements and SUCourse+

Students are responsible for all announcements made during the regular class meetings. Students should check their Sabanci University e-mail accounts (mymail.sabanciuniv.edu) and follow the SUCourse+ site (https://sucourse.sabanciuniv.edu/plus/) for this class regularly as they are responsible for all announcements and postings delivered through these sources.

Important Note: This document may be modified during the semester (Spring 2021-2022) due to unforeseen reasons.