CS512: Machine Learning
Syllabus, Fall 2020

Machine Learning is centered on automated methods that improve their own performance through learning patterns in data. In many data-rich domains, machine learning provides cost-effective solutions in business (image processing, speech recognition, recommendation and information retrieval systems) and in science (annotation of the genome, predicting disease biomarkers, etc). In this graduate-level class, students will get an introduction to the methodologies, technologies, and algorithms for machine learning. Topics include supervised learning, unsupervised learning, evaluating performance and model selection, a basic introduction to deep learning and reinforcement learning. The course will also discuss recent applications of machine learning. Programming and analytical assignments include hands-on practice with various learning algorithms, and a larger course project will give students an opportunity to work on a problem of their interest. Students entering the class are expected to have a pre-existing working knowledge of probability, statistics, linear algebra, programming, and algorithms.

Schedule
Needs some change. We shall discuss in class.

Contact Information
Instructors: Oznur Tastan, otastan@sabanciuniv.edu
Office Hours: Oznur Tastan, by appointment via e-mail and online.

Delivery format We will have synchronous lectures and recitations. Your active participation is expected. Generally, watching videos later takes more time and is more boring.

Course Webpage
We will be using SuCourse+. Please check regularly the SuCourse of the course for lecture notes, homework assignments, project information, discussions and announcements.

Textbook: No required textbooks. There will be required readings posted on SuCourse.

Optional textbooks
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R (online version available).

For further reading lists, please check here: https://machinelearningmastery.com/machine-learning-books/
Tentative Topics

Introduction
Basic concepts.
Probability and statistics review.
Model selection and feature selection.
Performance metrics.

Supervised learning
Naive bayes classifier.
Gaussian naive bayes classifier.
Linear classifiers: Logistic regression, Linear discriminant analysis
Linear regression.
Support vector machines.
Decision trees.
Ensemble methods: Bagging, boosting.
Neural networks. Basic introduction to deep learning.

Unsupervised Learning
Clustering Principal components analysis

Additional Topics (if time permits)
Hidden Markov Models
Reinforcement Learning
Active learning
Matrix factorization and collaborative filtering

Grading
The final grades will be based on the following:

• One exam (25%) that takes places at the 2/3 of the semester.
• One quiz, one-hour long, known date, at the end of the first month(10%).
• Three or four homework assignments, including programing (40%).
• One term project topic of your choice, deliverables include project proposal, progress and final reports and two presentations (progress and final) (25%).
• There may be oral exams to ensure a fair assessment in homework, quizzes and the exams. The student will be called for an exam in short notice. The instructor might ask questions from a broader range of topics to evaluate student’s understanding of the course material

IMPORTANT: One of the following conditions will result with an automatic F regardless of other grades:

1. Not submitting more than one assignment (empty homeworks do not count as a submission)
2. Average of the homework is below 30.
3. Not submitting a project report.
4. Being absent in a project presentation without a medical report.
5. Missing more the quiz or the exam without a medical report.
*** Not falling in one of the conditions does not guarantee passing the course, if your overall
performance is poor, you will fail the course.

**Homework submissions:** You may program in any programming language of your choice. I
recommend using Python as there are many useful libraries available. Your answers to a homework
assignment, including plots and mathematical work, should be submitted as a single PDF file.
Please follow the submission formats carefull, this course is understuffed in terms of TA. Your
cooperation is expected.

It is recommended that you use software that typesets mathematics (e.g. LATEX). . However, if
you need to you, you may scan your handwritten answers, convert to a PDF and submit. It should
be legible. Whenever you need to include source code in your LATEX document, you may find the
minted package convenient.

**Late day policy (IMPORTANT:** Each student will have a total of 4 free late (calendar) days
to use for homework assignments. You do not need to explain why you are submitting late and no
need to notify us. <= 24 hours late counts as 1 day late, etc. Once these total of four late days
are exhausted, any assignments turned in late will be penalized and will incur a reduction of 33%
in the final score, for each day (or part thereof) it is late. For example, if an assignment is up to
< 24 hours late, it incurs a penalty of 33%. Else if it is up to more than 24 hours and less than 48
hours late, it incurs a penalty of 66%. And if it is 72 or more hours late, it will receive no credit.

**Regrade policy:** Important: You may object a grade within 14 days after the grades are an-
nounced. If you feel that an error was made in grading your homework or midterm, please get an
appointment to discuss. Please note your grade may increase or decrease based on the evaluation.

**Honor code:** This course follows the Sabanci University Code of Academic Integrity, as explained
in the Student Disciplinary Rules and Regulation. Violations of the rules will not be tolerated.
Students may discuss and work on homework problems in groups in an abstract way. However,
each student must write down the solutions independently, and without referring to written notes
from the joint session. In other words, each student must understand the solution well enough in
order to reconstruct it by him/herself. In addition, each student must write on the problem set the
names of the people with whom s/he collaborated or discussed.

**Project**

The purpose of the project is to increase your knowledge about machine learning and get hands
on practical experience. Any project in the machine learning field that is feasible to accomplish in
the given time can be proposed. You will work groups of three people. The project can involve
applying known methods to solve an interesting question, or it can also involve coming up with a
new methodology to solve an existing problem on an existing data set.

You are responsible of proposing the data and the algorithm. The deliverables are (i) a proposal
write up, presentation, (ii) a progress report and presentation (iii) and final report and presentation.
The grade for the project will include a peer grade.

**Proposal Write Up:** Maximum one page (single spaced) proposal write up. Identify the data
set or data sets that you will be using. You should give a clear description of the characteristics
of the data (how many examples, what kinds of features do we have for each example, are there
issues with missing data or bad data, etc.). How will you evaluate performance? In certain settings,
you may want to try a few different performance measures. Identify a few "baseline algorithms". These are simple algorithms for solving the problem, such as always predicting the majority class for a classification problem, using a small set of decision rules designed by hand, or using a ridge regression model on a basic feature set. Ideally, you will be able to report the performance of a couple baseline algorithms in your proposal. The goal will be to beat the baseline, so if the baseline is already quite high, you will have a challenge. What methods do you plan to try to solve your problem, along with a rough timeline. Methods include data preprocessing, feature generation, and the ML models you’ll be trying. Once you start your investigation, it’s best to use an iterative approach, where the method you choose next is based on an understanding of the results of the previous step.

It should contain the following information: (1) project title, (2) team mates, (3) a high-level description of the problem you are trying to solve (4) The data sets you will be using. You should give description of the characteristics of the data (how many examples, what kinds of features do we have for each example, are there issues with missing data or bad data, etc.)(5) what you plan to achieve by the milestone.

**Progress Report:** The progress stage will significantly affect the final grade of the project. You are expected to have results by the progress date. At least three pages (single spaced). Include: (1) a high quality introduction and background information, (2) a clear description of the datasets you have used (3) methods you have used, data preprocessing, feature generation, and the ML models you have tried (4) the next steps, along with a rough timeline and (5) a clear description of the division of work among teammates.

**Progress Presentation:** There will be a 8 + 2 minute in class progress presentations on the work done so far.

**Final Project Presentation:** There will be a 8 + 5 minute in class final project presentation. Be prepared to answer questions not only what you have done but also on the details of the techniques.

**Final Report:** A final write up of the project. You should submit a pdf file electronically. It should have the following format:

- **Introduction:** A quick summary of the problem, methods and results.
- **Problem description:** Detailed description of the problem. What question are you trying to address?
- **Methods:** Description of methods and datasets used.
- **Results:** The results of applying the methods to the data set. Include the list of questions your experiments are designed to answer. Details of the experiments; observations.
- **Discussion:** Interpretation and discussion of the results.
- **Conclusions:** What is the answer to the question? What did you learn about the methods? Mention any future directions of interest.
- **Appendix:** A clear description of the contribution of each person. You may also include extra material (results, methods details) if needed in the appendix.

**Peer Grade:** When submitting the final project report, you are required to email your peer grades for each of your teammates. The peer grades will be 0 to 5. 5 being the highest grade. If you do not email, yourself will receive peer grade 0.
Disclaimer

- This syllabus and course details might need to be updated throughout the semester because of the uncertainties the pandemic brings. Any modification will be announced at SUCourse and also during the class. Students are responsible from following the announcements.
- The class material, including video lectures or sound recordings, must not be shared outside of class.
- Each of us is responsible for creating a safe and inclusive classroom experience for everyone in the class.