Course Syllabus Fall 2017

CS 464: Introduction to Machine Learning

Instructor: Mehmet Koyutürk[a]

[a]The material that will be used in this course is based on the material developed by Prof. Öznur Taştan.

Course Description

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 genome, predicting disease biomarkers, etc). In this undergraduate-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. The course will also discuss recent applications of machine learning. Programming and theoretical 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.

Learning Objectives

- Conceptual understanding of learning applications, formulation of learning tasks as computational problems, and methods that are designed to solve these problems
- Understanding the commonalities and differences between different learning tasks and different approaches to learning
- Understanding the trade-offs in developing solutions
- Thorough understanding of rigorously evaluating the performance of learning algorithms
- Ability to manipulate, extend, and apply machine learning methods and algorithms in the context of real-world problems

Prerequisites

CS-102, MATH-225 and MATH-230. Equivalent courses are accepted.

Class Meeting

Tue 16:40-17:30, Fri 13:40-15:30, EE-05. Spare Hour: Tue 15:40-16:30.

Course Personnel

Instructor: Mehmet Koyutürk, `koyuturk@cs.bilkent.edu.tr`
Office: EA-501, Office Hours: Tue 14:00-15:00, Wed 11:00-12:00

Teaching Assistant: Iman Deznabi, `iman.deznabi@bilkent.edu.tr`
Office: EA-427, Office Hour TBA

Course Webpage

We will be using Moodle. Please check regularly the Moodle page of the course for lecture notes, homework assignments, project information, discussions and announcements.

Textbook and Reading Material

No required textbooks. There will be **required readings** posted on Moodle.

Optional textbooks
Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2011.
Ethem Apaydin, Introduction to Machine Learning, 2e. The MIT Press, 2010. Kevin P. Murphy, Machine Learning: a Probabilistic Perspective, The MIT Press, 2012.
Tom Mitchell, Machine Learning, McGraw Hill, 1997.

Tentative Calendar

Introduction (≈ 3 weeks)
Basic concepts.
Model selection and feature selection.
Performance metrics and validation.
Probability and statistics review.

Supervised learning (≈7 weeks)
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.
Deep learning.

Unsupervised Learning (≈2 weeks)
Clustering: K-means, hierarchical clustering, DB-scan.
Principal components analysis.

Additional Topics (≈1 week)
Hidden Markov Models.
Active learning.

Coursework and Grading

The final grades will be based on the following:

- One exam (30%) that takes places at the 2/3 of the semester.
- Two quizzes (20%).
- Three homework assignments, including programing (25%).

- One term project topic of your choice, deliverables include three project proposal, progress, and final reports and presentations (25%). More information on the project is provided below.

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 don't 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.
5. Missing more than one quiz without a medical report.

*** Not falling in one of the conditions does not guarantee passing the course, if your overall performance or performance in the exams and quizes are poor, you will fail the course.

Late day policy: Each student will have a total of four free late (calendar) days to use for homeworks. Once these 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.

Homework regrade policy: If you feel that an error was made in grading your homework, please make an appointment with your TA and stop by in his office. Please note that regrading of a homework may cause your grade to go up or down. Important: You may object your homework within 14 days after the grades are announced.

Honor code: This course follows the Bilkent 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. 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.

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 in teams of five students. The project can involve applying known methods to solve an interesting question, or it can involve coming up with a new methodology to solve an existing problem on an existing data set.

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

Proposal Presentation (on Tue-Fri, Oct. 3-6): 5-minute presentation of the problem, approach, and validation plan, followed by 5-minute feedback from the instructors.

Proposal Document (due Mon., Oct. 9): Maximum one page (single spaced) proposal write up. It should contain the following information: (1) project title, (2) team mates, (3) description of the data, (3) precise description of the question you are trying to answer with the data, (4) comprehensive plan on how you will test your method. You must take into account the feedback provided during your presentation while preparing your proposal document.

Progress Presentation (on Wed, Nov. 7-10): There will be a 10 minute in class progress presentation

on the work done so far. During this presentation, you will receive feedback on how you can steer your project to ensure successful completion.

Progress Report (due Mon., Nov. 13): 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: (i) a high quality introduction and background information, (ii) what have you done so far, you are expected to be in the implementation stage, (iii) what remains to be done and (iv) a clear description of the division of work among teammates.

Final Report (due Fri., Dec 15): 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, please e-mail your TA with peer grades for each of your teammates. The peer grades will be 0 to 5, 5 being the highest grade.

Final Project Presentation (on Tue-Fri, Dec 19-22): There will be a 10 + 2 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.