CS464: Introduction to Machine Learning

Syllabus, Spring 2015

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 (speech recognition systems, recommendation systems, 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 (generative/discriminative learning, parametric/non-parametric learning), unsupervised learning (clustering, dimensionality reduction), 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.

Schedule

Lectures: Tue 1:40 – 3:30; Thu: 15:40 – 16:40 (spare hour: Thu 16:40 – 17:30, we will resort to this hour whenever necessary), BZ05
Office Hours: Mustafa Buyukozkan EA 425 To be announced; Oznur Tastan EA 429, by appointment through email.

Contact Information

Instructor: Oznur Tastan, `oznur.tastan@cs.bilkent.edu.tr`
Teaching Assistant: Mustafa Buyukozkan, `m.buyukozkan@bilkent.edu.tr`
Graders: Eren Golge, `eren.golge@cs.bilkent.edu.tr`
Burak Uzun `burak.uzun@bilkent.edu.tr`

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

Optional textbooks
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.
Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2011.

Tentative Syllabus

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

Supervised learning
Naive Bayes Classifier.
Gausian Naive Bayes Classifier.
Linear classifiers: Logistic regression, LDA, perceptron
Linear regression.
Support vector machines.
Decision trees.
Ensemble methods: Bagging, boosting.

Unsupervised Learning
Clustering: K-means, hiereachical clustering.
Principal components analysis.

Additional Topics
Active learning.
Hidden Markov Models.

Grading

The final grades will be based on the following:

- One exam (30%) that takes places at the 2/3 of the semester.
- Three quizzes (20%).
- Four (tentative) homework assignments, including programing (25%).
- One term project topic of your choice, deliverables include three reports and three presentations (25%).

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.

The conditions above are minimum requirements, not falling in one of the conditions does not guarantee passing the course.

Late day policy: Each student will have 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 stop by to your TA's office hour. Please note that regrading of a homework may cause your grade to go up or down.

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 should 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 groups of two or 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. 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) what you plan to achieve by the milestone.

Proposal Presentation: There will be a 3 minute in-class presentation.

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: (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.

Progress Presentation: There will be a $10 + 2$ minute in class progress presentations on the work done so far.

Final Project Presentation: There will be a $10 + 5$ 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 datsets used.
- Results: The results of applying the methods to the data set. Inlcude 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.