Syllabus: CS 4269/6362 Machine Learning
TR, 11-12:15pm

Instructor: Yevgeniy (Eugene) Vorobeychik
Jacobs Hall, 379
yevgeniy.vorobeychik@vanderbilt.edu
OFFICE HOURS: TR 12:15-1:15pm, or by appointment

TA: Sweta Panda

Spring 2016

1 Introduction

This course takes an application driven approach to current topics in machine learning. The course covers supervised learning (classification/regression) and unsupervised learning (dimensionality reduction, clustering). Additional topics may include reinforcement learning and learning theory. The course will also consider challenges resulting from learning applications. We will cover popular algorithms (Naive Bayes, SVM, perceptron, HMM, k-means, maximum entropy) and will focus on how statistical learning algorithms are applied to real world applications. Students in the course will implement several learning algorithms and develop a learning system for a final project.

2 Course Materials and Prerequisites

The prerequisites for this course are CS 260, CS 360, or equivalent. In addition, a strong background in probability and statistics, as well as linear algebra, will be assumed. Finally, homework assignments will involve Java programming, so previous experience with Java is strongly recommended.

We will be using the following textbook in the class:

Chris Bishop. Pattern Recognition and Machine Learning.

3 Grading

3.1 CS 269

Grading for the CS 4269 course is as follows:
Homeworks: 50%
Team Project: 40%
Class participation: 10%

3.2 CS 362

Grading for the CS 6362 course is as follows:

Homeworks: 50%
Presentation: 10%
Individual Project: 30%
Class participation: 10%

3.3 Details

Homeworks: There will be 5 homework assignments, based on the material covered in lectures. Homeworks will involve significant Java programming.

Presentation: Each CS 6362 (graduate course) student will present a paper. 25 minutes will be allocated for this presentation, which should cover the relevant background, in addition to describing the results in the paper itself.

Class participation: Students will be expected to regularly attend class and actively participate. In particular, the expectation is that all of the papers presented by peers in the second half of the course will be read by all students in advance, and students should come to class prepared to engage in thoughtful discussion.

Course project: The course project will entail developing a complex machine learning system. This can be either a research project, or applied research to an interesting and practically relevant domain. Here is the schedule for the course project:

- CS 4269 teams should form during the first week of the course (not applicable to CS 6362 students), and should choose the project topic by the end of week 2 (1/22/2016), at which point they will need to submit a preliminary project plan, which will make up 5% of the grade (a list of possible topics can be found at the end of the syllabus).
- After nine weeks, students will need to submit a project progress report, which will report on significant progress made on the project by that point. This progress report will account for another 10% of the grade for CS 4269 students, and 5% of the grade for CS 6362 students.
- The final project report, due at the end of the course, is worth the remaining portion of the project grade allocated to the course project (25% of the grade for CS 4269 students and 20% for CS 6362 students). Final project presentation (in class) will account for 5% of this. For CS 4269 students, participation (a poster presentation) in Design Day is also required, and constitutes 5% of the final grade.

4 Course Schedule

Week 1 (1/12 and 1/14): Introduction and Review

1. Introduction [**Ch. 1-2**]
2. Math review
3. Probability and statistics review
4. **Homework 1: fundamentals of ML [due: 1/19]**
5. **CS 4269: teammate preferences submitted [due: 1/14]**

I. SUPERVISED LEARNING

Week 2 (1/19 and 1/21): Linear Supervised Learning

- Linear Regression [**Ch. 3**]
- Bias-Variance decomposition [**Section 3.2**]
- Linear Classification, perceptron algorithm, logistic regression, Naive Bayes [**Ch. 4**]
- **Homework 2: linear supervised learning [due: 1/28]**
- **Preliminary project plan [due: 1/22]**

Week 3-4 (1/26, 1/28, 2/2, 2/4): Non-Linear Supervised Learning

- Decision Trees [**Section 14.4**]
- Neural Networks [**Ch. 5**]
- Support Vector Machines [**Ch. 7, Section 7.1**]
- Kernel Methods [**Ch. 6**]
- Nearest-Neighbor Methods [**Section 2.5.2**]
- Ensemble Methods, Boosting [**Ch. 14**]
- **Homework 3: non-linear supervised learning [due: 2/11]**

Week 5 (2/9, 2/11): Regularization and Dimensionality Reduction

- Regularization (L_p-regularization, ridge regression, Lasso) [**Section 3.1.4**]
- Dimensionality reduction: PCA [**Ch. 12**]

II. UNSUPERVISED LEARNING

Week 6 (2/16, 2/18): Clustering

- Clustering [**Section 9.1**]
- Density estimation [**Section 2.5.1**]
- Gaussian mixture models [**Section 9.2**]
- EM algorithm [**Section 9.4**]
- **Homework 4: Dimensionality reduction and unsupervised learning [due: 2/25]**
- **2/16:** Invited Lecture on Deep Learning (Professor Thomas Lasko, biometical informatics)

III. GRAPHICAL MODELS

Week 7-8 (2/23, 2/25, 3/2, 3/4): Graphical Models

- Bayes Networks [**Section 8.1-8.2**]
- Markov Random Fields [**Section 8.3**]
- Inference [**Section 8.4**]
- Sequential graphical models (dynamic Bayes networks, HMMs) [**Ch. 13**]
- **2/27:** Invited Lecture on Detecting Misuse of Electronic Medical Records (Professor Daniel Fabbri, biometical informatics)
- **Homework 5: graphical models [due: 3/15]**
- **Project progress report [due: 3/15]**

IV. REINFORCEMENT LEARNING (if time allows)

Week 9 (3/15, 3/17): Reinforcement Learning

- Markov decision processes, value and policy iteration
- Multi-armed bandit
- Q-Learning, temporal difference learning

V. Presentations: 3/22, 3/24, 3/29, 3/31, 4/5, 4/7, 4/12, 4/14

VI. Final project presentations: 4/19, 4/21

5 Some Possible Course Project Topics

- Stock price prediction
- Predicting protein interactions
- Predicting future GDP / unemployment rate, especially as a function of policy decisions
- Spam/phishing email filtering
- Predicting market effects of hospital mergers
- Predicting auction prices on eBay
- Predicting (pro or college) basketball/football/baseball final scores (and winner)
- Predicting dynamic behavior of social, physical, and economic networks (i.e., edge addition and removal)
- Trading Agent Competition
- Predicting future task/process characteristics in operating systems
- Predicting CPU/Hard Drive utilization