MACS 30133 - Machine Learning for Political Analysis

Computational Social Science - Division of the Social Sciences
University of Chicago - **Spring/2020**

Instructor: Dr. Diogo Ferrari
e-mail: diogoferrari@uchicago.edu
Office: 1155 E. 60th St. Room 221
Office-hours: Monday 15:00-17:00 (or by appointment)

Meeting day/time MWF 10:30-11:20
Location: Public Policy 140B

Overview

This is an intermediate-to-advanced introduction to the mathematical and computational aspects of the core statistical and machine learning techniques. The goal is to equip students with a knowledge of the theoretical and practical aspects of four groups of machine learning methods which are widely used in applied research: (1) dimension reduction (PCA, MDS, and their extensions) (2) classification methods (SVM, Bayes classifiers, and other classification methods) (3) clustering procedures and density estimation (K-means, FMM, non- and semi-parametric Bayesian methods) (4) categorical data analysis (with brief introduction to probabilistic graphical models). The course includes applications in Political Science, such as FMM to estimate fraud in elections, PCA to construct indices to measure democracy, and text classification.

Prerequisite: proficiency in R or Python; basic calculus; probability and statistics (regression, expectation), basic linear algebra

Objectives

In this course, students will:

1. Develop a rigorous understanding of widely used machine learning methods
2. Develop the ability to extend existing approaches, and adapt them for novel and different problems
3. Develop the ability to critically evaluate the adequacy of machine learning methods and their limitations

Course structure

The course is divided into four parts, which one covering a group of machine learning methods:

1. Dimension reduction methods
2. Classification methods
3. Clustering procedures
4. Graphical models for multivariate and categorical data

Evaluation

Evaluation

Component	Number	Weight
Assignments	5	40%
Final Presentation	1	20%
Final Project	1	40%

Grading Scale

Quality Performance	Letter Grade	Points min	max	Round interval
Excellent	A	93	100	[92.5, 100]
	A-	90	92	[89.5, 92.5)
Good	B+	87	89	[86.5, 89.5)
	B	83	86	[82.5, 86.5)
	B-	80	82	[79.5, 82.5)
Satisfactory	C+	77	79	[76.5, 79.5)
	C	70	76	[69.5, 76.5)
Unsatisfactory	D	51	69	[50, 69.5)
Failure	F	0	50	[0, 50)

Rounding If needed, the point scale will be rounded using the round half up rule. It means, for instance, that any grade $x \in [92.5, 93.5)$ becomes 93, and a grade $x \in [91.5, 92.5)$ becomes 92. An exception applies to the bottom of the scale, as indicated in the table above.

Textbook and Materials

The lectures will be based on the following textbooks and lecture notes:

- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. Springer (available online at https://web.stanford.edu/~hastie/ElemStatLearn/)
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* MIT press (information available at https://www.cs.ubc.ca/~murphyk/MLbook/)
- Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer
- Bishop, C. M. et al. (1995). *Neural networks for pattern recognition.* Oxford university press
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning.* Springer (available at http://faculty.marshall.usc.edu/gareth-james/ISL/)
- Smola, A. and Vishwanathan, S. (2008). Introduction to machine learning. *Cambridge University, UK*, 32:34 (available at http://alex.smola.org/drafts/thebook.pdf)

Diversity Statement

This course is open to all students who meet the academic requirements for participation. Any student who has a documented need for accommodation should contact Student Disability Services (773-702-6000 or disabilities@uchicago.edu) and the instructor as soon as possible.

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity students bring to this class be viewed as a resource, strength, and benefit. Please let me know ways to improve the effectiveness of the course for you personally, or for other students or student groups. Your suggestions are encouraged and appreciated.

It is my intent to present materials and activities that are respectful of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture. I will attempt to foster an environment in which each class member is able to hear and respect one another. It is my intent to maintain an atmosphere of trust and safety in the classroom. Please let me know if something said or done in the classroom, by either myself or other students, is particularly troubling or causes discomfort or offense. While our intention may not be to cause discomfort or offense, the impact of what happens throughout the course is not to be ignored and is something that I consider to be very important and deserving of attention. If and when this occurs, there are several ways to alleviate some of the discomfort or hurt you may experience:

1. Discuss the situation privately with me. I am always open to listening to students' experiences and want to work with students to find acceptable ways to process and address the issue.

2. Discuss the situation with the class. Chances are there is at least one other student in the class who had a similar response to the material. Discussion enhances the ability for all class participants to have a fuller understanding of context and impact of course material and class discussions.

3. Notify me of the issue through another source such as your preceptor, a trusted faculty member, or a peer. If for any reason you do not feel comfortable discussing the issue directly with me, I encourage you to contact your preceptor and/or your program's Diversity and Inclusion representative: Darcy Heuring (MAPSS), Matthias Staisch (CIR), and Chad Cyrenne (Computation). You are also welcome and encouraged to contact the Faculty Director of your program.

The University of Chicago is committed to diversity and rigorous inquiry from multiple perspectives. The MAPSS, CIR, and Computation programs share this commitment and seek to foster productive learning environments based upon inclusion, open communication, and mutual respect for a diverse range of identities, experiences, and positions. Any suggestions for how we might further such objectives both in and outside the classroom are appreciated and will be given serious consideration. Please share your suggestions or concerns with your instructor, your preceptor, or your program's Diversity and Inclusion representatives: Darcy Heuring (MAPSS), Matthias Staisch (CIR), and Chad Cyrenne (Computation). You are also welcome and encouraged to contact the Faculty Director of your program.

Policy on academic honesty

The University of Chicago has a formal policy on academic honesty that you are expected to adhere to. Here are some guidelines we expect you to follow:

1. Courtesy, honesty, and respect should be shown by students toward faculty members, guest lecturers, administrative support staff, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the School.

2. Academic dishonesty can encompass many activities, which includes plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. One of the gravest academic dishonesty is plagiarism: knowingly handing in someone else's work as your own, whether it be work done by another student in the class or available publicly on the Internet.

3. The preparation of solutions for problem sets, papers, and examinations, assigned on an individual basis, must represent each students own effort. Therefore:
   - You MUST NOT copy or use someone else's work (with or without their permission) in your own solution. You have to write your own.
   - DO NOT post your solutions to problem sets or exams in publicly-accessible websites, like pastebin, a public GitHub repository, GitHub gists, etc. While these tools may seem like convenient mechanisms for sharing code with an instructor/TA or with a project partner, they can also expose your code to other students in the class. If you do post your solution in a publicly-accessible location, and we find out about it outside of a plagiarism incident, you will just get a warning. However, if another student in the class uses code that you posted on such a site (even if you did not intend for that code to be used by someone else), you be considered an equally guilty party in a plagiarism offense, and will receive the exact same penalty as the student who used your code.

References

Bishop, C. M. (2006). *Pattern recognition and machine learning.* springer.

Bishop, C. M. et al. (1995). *Neural networks for pattern recognition.* Oxford university press.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. Springer.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning.* Springer.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective.* MIT press.

Smola, A. and Vishwanathan, S. (2008). Introduction to machine learning. *Cambridge University, UK*, 32:34.