IS 698-05 / 800-02 Special Topics in Information Systems: Probabilistic Machine Learning (Fall 2018)

Information Systems Department
University of Maryland Baltimore County
Baltimore, Maryland 21250
Departmental Office: Room ITE 404, ph. 410-455-3206

Course Description

In this course, students will develop an understanding of probabilistic machine learning techniques for data analysis, and how to apply these methods in practice. Probabilistic methods, which provide a principled foundation for reasoning under uncertainty, underpin many modern machine learning methods for a broad spectrum of application domains, from text analytics, to recommender systems, to bioinformatics, and more. The course will cover modeling techniques such as directed probabilistic graphical models (a.k.a. Bayesian networks), parameter estimation methods including maximum likelihood estimation and Bayesian inference, and their application to areas such as natural language processing.

Student learning outcomes: By the end of this course, you will be able to:

- Apply probabilistic machine learning methods to a variety of problem domains,
- navigate the practical challenges in using these methods, including evaluation and convergence assessment,
- and discuss the underlying theoretical principles behind them.

Lecture time and venue: Tuesdays 4:30pm - 7:00pm Sherman Hall 210

Instructor: Dr. James Foulds
Instructor email: jfoulds [at] umbc [dot] edu. Please use Piazza for course-related questions, instead of email, so that everyone can benefit from the answers.
Instructor office hours: Tuesdays 3 - 4pm ITE 447 (other times by appointment)

Piazza: Sign up for this course at piazza.com/umbc/fall2018/is6980580002
Poll Everywhere: Vote on in-class poll questions at PollEv.com/jamesfoulds656 . Register your account for the course at polleverywhere.com/register?p=6bu49-6kgn&u=wJdvD6sW , by week 2 in order to get participation credits.

Prerequisites

- Requires IS733, or IS777, or consent of the instructor.
- Students will need undergraduate-level knowledge of probability, linear algebra, elementary calculus, and basic programming ability in a high-level language such as Java, Python, R, or Matlab. The course will begin with a review of probability theory, but students are encouraged to brush up on this beforehand, using e.g. the book "A First Course in Probability" by Sheldon Ross.

Required Textbook

Modeling and Reasoning with Bayesian Networks, by Adnan Darwiche.

We will also sometimes use Information Theory, Inference, and Learning Algorithms (Mackay, 2003), which is available as a free PDF download at the author's webpage.

Course Requirements and Grading

- Homeworks 25% (5 of them, 6.25% each for your best 4 homeworks; the lowest score will be dropped)
- Group projects 35%:
  - Proposal 5% (due 10/2/2018)
  - Mid-term report 5% (due 11/6/2018)
  - Group project poster 10% (presented in class 12/11/2018, digital copy due at the same time)
  - Final report 15% (due **Thursday** 12/13/2018, 11:59pm, by email)
- Midterm 10% (10/23/2018)
- Final 25% (12/18/2018)
- Participation 5%
  - Poll questions 4%
  - At least two Piazza posts 1% (can be either questions or answers)

The project will be done in groups of 2-3 students. Project proposals are to be sent to me by email, and approved by the deadline.

In this course, participation means more than just showing up. It also refers to contributing to everyone's learning, through active engagement in peer instruction exercises, in-class discussions, and Piazza questions/answers. Participation grades will be assessed as a percentage of peer instruction questions answered (correctly or not), with a 90% response rate being sufficient for full points, and by Piazza contributions. Two or more contributions (either questions or answers) on Piazza will earn you 1% of the final grade.

With respect to final letter grades, UMBC's Catalog states that an A indicates "superior" achievement; B, "good" performance; C, "not satisfactory"; D, "unacceptable"; F, "failure." There is specifically no mention of any numerical scores associated with these letter grades. Below is how grades may be assigned based on your final points, accumulated over the semester. Grades will be assigned using a plus/minus system. It is university policy that A+, D+, and D- are not assigned. I do not grade on a curve, so that everyone in the class has the opportunity to succeed.

Final Grade	Letter Grade	Points when calculating GPA
91 - 100	A	4.0
89 - 90.99	A-	3.7
87 - 88.99	B+	3.3
81 – 86.99	B	3.0
79 - 80.99	B-	2.7
77 – 78.99	C+	2.3
71 – 76.99	C	2.0
69 - 70.99	C-	1.7
60 – 68.99	D	1.0

0 – 59.99	F	0.0

Homework and Exam Policies

- Homeworks are due at the beginning of class on the dates specified. Late homeworks will not be accepted unless an extension is approved by me in advance. Requests for extensions must be made at least three days before the due date.
- In the event of class cancellation due to inclement weather, any hard-copy paper assignment or test will be due in the next class meeting. Electronic submissions will still be due on the original due date.

Schedule

Lecture		Summary	Details	Assessment	Required reading
9/4/2018	Week 1	Course overview	Intro to probabilistic modeling. Motivating examples. Box's loop		Bishop, C. M. (2013). Model-based machine learning. Phil. Trans. R. Soc. A, 371(1984).
9/11/2018	Week 2	Probability	Random variables, conditional and joint probabilities, independence, Bayes rule	**HW1 out**	Information Theory, Inference, and Learning Algorithms (Mackay, 2003), Ch. 2
9/18/2018	Week 3	Bayesian networks	Bayesian networks, d-separation, Markov blankets		Darwiche ch. 4
9/25/2018	Week 4	Modeling with Bayesian networks	MPE, MAP, modeling examples. Project brainstorming.	**HW1 due, HW2 out. Project groups formed by this date**	Darwiche ch. 5
10/2/2018	Week 5	Inference in Bayesian networks	Variable elimination, bucket elimination	**Project proposal due**	Darwiche ch. 6
10/9/2018	Week 6	Maximum likelihood estimation	Finding MLE for simple models	**HW2 due, HW3 out**	Darwiche ch. 17
10/16/2018	Week 7	Bayesian inference	frequentist vs Bayesian inference, MAP vs MLE, full posterior vs point		Bishop, C. M. (2013). Model-based machine learning. Phil. Trans. R. Soc.

			estimates, posterior predictive		A, 371(1984)., Ch. 3
10/23/2018	Week 8		Covers up to homework 2	**Midterm exam**	
10/30/2018	Week 9	Generative models for discrete data	Conjugate priors, beta/Bernoulli, Dirichlet/multinomial, urn process interpretations, naive Bayes document model	**HW3 due, HW4 out**	Darwiche, ch. 18.4
11/6/2018	Week 10	Mixture models and hidden Markov models	Mixtures of Gaussians/connection to K-means, mixtures of multinomials, HMMs, Bayesian inference for HMMs	**Project mid-term progress report due**	Witten and Frank (2017), Data Mining (4th Edition), Ch 9.3 (up to Clustering with Correlated Attributes), Ch. 9.8 (up to Hidden Markov Models) (if you have it from IS733, otherwise Wikipedia articles Mixture Model, Hidden Markov Model)
11/13/2018	Week 11	Topic models and mixed membership models	LSA/PLSA, Genetic admixtures, LDA, collapsed Gibbs sampler	**HW4 due, HW5 out**	Blei, David M. (2012). Probabilistic topic models. Communications of the ACM, 55(4), 77-84
11/20/2018	Week 12	Evaluating unsupervised models	Log-likelihood on held-out data, posterior predictive checks, correlation with metadata, human evaluation		Dave Blei's notes on posterior predictive checks
11/27/2018	Week 13	Social network models	Erdos-Renyi random graphs, stochastic blockmodels, MMSB, latent space models	**HW5 due**	Goldenberg A., Zheng, A.X., Fienberg, S.E. and Airoldi, E.M. (2010). A Survey of Statistical Network Models. Foundations and Trends® in Machine Learning: Vol. 2:

					No. 2 Ch. 3, can skip 3.3 and 3.7
12/4/2018	Week 14	Models for computational biology	Profile HMMs for protein sequence alignment. Phylogenetic models and coalescent models		What is Phylogeny? article from the Tree of Life Web Project, hosted at the University of Arizona. Teh, Y. W., and D. M. Roy. (2007). Bayesian Agglomerative Clustering with Coalescents. NIPS up to Section 2.
12/11/2018	Week 15	Group project presentations		**Digital copies of posters due. Project final report due TODO (11:59pm)**	
12/18/2018	Exam week			**Final exam 3:30-5:30pm (Sherman Hall 210)**	

The schedule may be subject to change. The summary and details columns are only a guideline of the content likely to be covered, and the dates on which material is covered may shift.

Instructional Methods

Traditional lectures will be augmented with active learning methods, primarily in the form of peer instruction exercises. Research has strongly indicated that active learning improves student outcomes in STEM fields versus traditional lecturing (Freeman et al., 2013). We will be using the Poll Everywhere service for polls and quizzes. You will need to bring a mobile device, laptop, or tablet to class in order to participate in the exercises. If you do not have a suitable device, please let me know as soon as possible.

Pre-class reading assignments will be given for each lesson, which are very important for learning, and for making the best use of our limited time together (a partially "flipped classroom" approach). These readings are therefore required.

Software

This course will make extensive use of the PyMC3 python package for probabilistic machine learning.

Academic Integrity

UMBC's policies on academic integrity will be strictly enforced (see the University System of Maryland's policy document, UMBC's academic integrity overview page, the student academic conduct policy and the UMBC catalog). In particular, all of your work must be your own. Acknowledge and cite source material in your papers or assignments. While you may verbally discuss assignments with your peers, you may not copy or look at anyone else's written assignment work or code, or share your own solutions. Any exceptions will result in a zero on the assessment in question, and may lead to further disciplinary action. Some relevant excerpts from UMBC's policies, as linked to above, are:

- "By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC's scholarly community in which everyone's academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping others to commit these acts are all forms of academic dishonesty, and they are wrong." (UMBC's academic integrity overview)
- "Students shall not submit as their own work any work which has been prepared by others." (USM policy document)
- "Students shall refrain from acts of cheating and plagiarism or other acts of academic dishonesty." (USM policy document)
- "Plagiarism means knowingly, or by carelessness or negligence, representing as one's own, in any academic exercise, the intellectual or creative work of someone else." (student academic conduct policy)
- "Cheating means using or attempting to use unauthorized material, information, study aids, or another person's work in any academic exercise" (student academic conduct policy)

Accessibility in the Classroom; Student Support / Disability Services

UMBC is committed to eliminating discriminatory obstacles that may disadvantage students based on disability. Student Support Services (SSS) is the UMBC department designated to:

- receive and maintain confidential files of disability-related documentation,
- certify eligibility for services,
- determine reasonable accommodations,
- develop with each student plans for the provision of such accommodations, and
- serve as a liaison between faculty members and students regarding disability-related issues.

If you have a disability and want to request accommodations, contact SSS in the Math/Psych Building, Room 213 (or call 410-455-2459). SSS will require you to provide appropriate documentation of disability and complete a Request for Services form available at my.umbc.edu/groups/sss. If you require accommodations for this class, please make an appointment to meet with me to discuss your SSS-approved accommodations, so that we can best accommodate your needs in a confidential and timely manner.

Counseling Center

Diminished mental health can interfere with optimal academic performance. The source of symptoms might be related to your course work; if so, please speak with me. However, problems with other parts of your life can also contribute to decreased academic performance. UMBC provides cost-free and confidential mental health services through the Counseling Center to help you manage personal challenges that threaten your personal or academic well-being.

Remember, getting help is a smart and courageous thing to do -- for yourself and for those who care about you. For more resources get the Just in Case mental health resources Mobile and Web

App. This app can be accessed by clicking: counseling.umbc.edu/justincase.

The UMBC Counseling Center is in the Student Development & Success Center (between Chesapeake and Susquehanna Halls). Phone: 410-455-2472. Hours: Monday-Friday 8:30am-5:00pm.

Diversity Statement on Respect

Students in this class are encouraged to speak up and participate during our meetings. Because the class will represent a diversity of individual beliefs, backgrounds, and experiences, every member of this class must show respect for every other member of this class. (Statement from California State University, Chico's Office of Diversity and Inclusion).

Family Educational Rights and Privacy Act (FERPA) Notice

Please note that as per federal law I am unable to discuss grades over email. If you wish to discuss grades, please come to my office hours.

Campus Resources

- Diversity and inclusion resources: about.umbc.edu/diversity-and-inclusion/
- The Mosaic Center for Culture and Diversity: osl.umbc.edu/diversity/mosaic
- Career Center's resources for diverse populations (including student organizations): http://careers.umbc.edu/students/resources/diverse/
- Resources for LGBTQ students: osl.umbc.edu/lgbtq/community_resources/
- Office of International Education Services (IES): ies.umbc.edu/
- Information regarding recent executive actions: ies.umbc.edu/executive-actions/
- Wellness Initiative: wellness.umbc.edu/
- Counseling Center: counseling.umbc.edu/
- Women's Center: womenscenter.umbc.edu/
- Center for Women in Technology (CWIT): cwit.umbc.edu/
- Women Involved in Learning and Leadership (WILL) Program: gwst.umbc.edu/will/
- Sexual assault and relationship violence on-campus resources: womenscenter.umbc.edu/sexual-assault-and-relationship-violence-response-team-and-umbcs-voices-against-violence/
- Sexual misconduct policies and procedures (including filing a complaint): humanrelations.umbc.edu/sexual-misconduct/policies-and-procedures/
- University System of Maryland's Policy of Non-Discrimination on the Basis of Sexual Orientation and Gender Identity or Expression: humanrelations.umbc.edu/files/2014/12/USMPolicyNonDiscrimSOrientGenderIEJune2012.pdf
- Office of Student Disability Services: sds.umbc.edu/
- Academic Center for Student Athletes: umbcretrievers.com/information/academiccenter/acsa
- Veteran Services veterans.umbc.edu/
- The Interfaith Center: osl.umbc.edu/diversity/interfaith/
- Graduate Student Association: gsa.umbc.edu/
- Graduate Student Association Writing Advisor: gsa.umbc.edu/writing-advisor/