Advanced Topics on Machine Learning
Spring 2024

COURSE & CONTACT INFORMATION

Credits: 3 credits
Meeting Day(s): Monday and Wednesday
Meeting Time: 9:45am – 11:00am
Meeting Place: Health Sciences Edu Ctr 2-132

Instructor: Wei Pan, PhD
Email: panxx014@umn.edu
Office Hours: Monday and Wednesday, 11:00am-11:30am, Mayo D378 (during the weeks as the instructor)
Office Location: University Office Plaza, 2221 University Avenue SE, Room 235, Minneapolis

Instructor: Xiaotong Shen, PhD
Email: xshen@umn.edu
Office Phone: 612-624-7098
Office Hours: Monday and Wednesday, 11:00 am-12:00 pm (during the weeks as the instructor)
Office Location: Ford 391

TA: Michael Anderson
Email: and02709@umn.edu
Office Hours: Tuesday and Friday, 3:00pm-4:00pm
Office Location: University Office Plaza, 2221 University Avenue SE, Room TBA, or zoom https://umn-private.zoom.us/j/7013228968

COURSE DESCRIPTION

It covers a range of emerging topics in machine learning, data analytics, and big data. This course will introduce various statistical and computational techniques for prediction and inference. These techniques are directly applicable to Business Analytics, Engineering, and Biomedical Research. This course requires basic knowledge of machine learning and data mining (e.g., Pubh7475, Stat8053 or a similar course). Topics include the following:

- Data exploration and data science;
- Optimization for machine learning;
- High-dimensional analysis: prediction and inference;
- Deep neural network learning: basics (FNN, CNN, RNN/LSTM); advanced topics (Transformers, Diffusion models, etc);
- Recommender systems: personalized prediction;
- Undirected and directed graphical models;
- Unstructured data and text mining: Numerical embedding and language models;
- Causal Machine Learning.


Please visit regularly.

COURSE PREREQUISITES
Pubh7475 or Stat8053 or a similar course, or permission of instructor; familiarity with programming in R or Python.

COURSE GOALS & OBJECTIVES
After taking the course, the student should have a working knowledge of using some state-of-the-art of AI/M techniques and learning other emerging ones in practice.
METHODS OF INSTRUCTION AND WORK EXPECTATIONS

In-class lectures are the main method of instruction. Students are expected to come to class for active learning, e.g. participating in discussions, doing (reading and written) assignments, and (co-) writing a report and presenting for a course project towards the end of the semester. Late assignments or project reports are not accepted unless with legitimate reasons or advance permission from the instructor.

This is a 3-credit course. The University expects that for each credit, you will spend a minimum of three hours per week attending class or comparable online activity, reading, studying, completing assignments, etc. over a 15-week term. Thus, this course requires approximately $3 \times 45$ hours of effort spread over the term to earn an average grade.

In this course, students are expected to engage with each other in respectful and thoughtful ways.

COURSE TEXT & READINGS

No textbook. Slides and published research papers will be shared. Following is a list of suggested (optional) readings, which will be updated as the course progresses during the semester.

1. Introduction
   1) McKinsey Global Institute June 2011 Big data: The next frontier for innovation, competition, and productivity.
   2) Donoho D. (2015), 50 Years of Data Science. JCGS.
   4) Hand, D.J. (2006), Classifier Technology and the Illusion of Progress (with comments and a rejoinder by the author). Statist. Sci. 21, Iss. 1, 1-34.
   5) S. Guha, R. Hafen, J. Xia, J. Rounds, J. Li, B. Xi, and W. S. Cleveland (2012), Large complex data: divide and recombine (D&R) with RHIPE, Stat 1, 53-67.

2. Optimization

3. High-dimensional Analysis


5. Recommender systems

6. Graphical Models

7. Causal ML

8. Semi-supervised learning:

9. Network analysis:
1) Neuman MEJ. Detecting community structure in networks.

**COURSE OUTLINE/WEEKLY (TENTATIVE) SCHEDULE**

<table>
<thead>
<tr>
<th>Week</th>
<th>Topic</th>
<th>Readings</th>
<th>Activities/Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1, Jan 15-19</td>
<td>Data exploration and data science (Shen)</td>
<td>Selected papers in Group 1.</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 2, Jan 22-26</td>
<td>Optimization for machine learning (Shen)</td>
<td>Papers in Group 2</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 3, Jan 29-Feb 2</td>
<td>High-dimensional analysis (Pan)</td>
<td>Selected papers in Group 3.</td>
<td>Assignment 1, Readings</td>
</tr>
<tr>
<td>Week 4, Feb 5-9</td>
<td>Deep learning Basics: FNN, CNN (Pan)</td>
<td>Selected papers in Group 4</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 5, Feb 12-16</td>
<td>Python tutorial: Deep Learning (Guest: Liu)</td>
<td>Selected papers in Group 4</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 6, Feb 19-23</td>
<td>DL Advanced: RNN, LSTM, Models (Shen)</td>
<td>Selected papers in Group 4</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 7, Feb 26-Mar 1</td>
<td>Natural Language Processing and models (Shen)</td>
<td>Selected papers in Group 4</td>
<td>Assignment 2, Readings</td>
</tr>
<tr>
<td>Week 8, Mar 4-8</td>
<td>Spring Break (No Class)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 9, Mar 11-15</td>
<td>Undirected and directed graphical models (Pan)</td>
<td>Selected papers in Group 6</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 11, Mar 18-22</td>
<td>Diffusion Models, Normalizing Flows, Synthetic Data, and Applications (Shen)</td>
<td>Papers in Group 4</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 12, Mar 25-29</td>
<td>Recommender systems (Shen)</td>
<td>Selected papers in Group 5</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 13, Apr 1-5</td>
<td>Causal ML (Pan)</td>
<td>Selected papers in Group 7</td>
<td>Assignment 3, Readings</td>
</tr>
<tr>
<td>Week 13, Apr 8-12</td>
<td>Semi-supervised learning (Pan)</td>
<td>Selected papers in Group 8</td>
<td>Readings</td>
</tr>
<tr>
<td>Week 14, Apr 15-19</td>
<td>Student presentations</td>
<td></td>
<td>Readings</td>
</tr>
<tr>
<td>Week 15, Apr 22-26</td>
<td>Student presentations</td>
<td></td>
<td>Critiques</td>
</tr>
<tr>
<td>Week 16, Apr 29</td>
<td>Student Presentations</td>
<td></td>
<td>Critiques</td>
</tr>
</tbody>
</table>
LAND ACKNOWLEDGEMENT

The School of Public Health at the University of Minnesota Twin Cities is built within the traditional homelands of the Dakota people. Minnesota comes from the Dakota name for this region, Mni Sóta Maḳoce, which loosely translates to the land where the waters reflect the skies.

It is important to acknowledge the peoples on whose land we live, learn, and work as we seek to improve and strengthen our relations with our tribal nations. We also acknowledge that words are not enough. We must ensure that our institution provides support, resources, and programs that increase access to all aspects of higher education for our American Indian students, staff, faculty, and community members.

SPH AND UNIVERSITY POLICIES & RESOURCES

The School of Public Health maintains up-to-date information about resources available to students, as well as formal course policies, on our website at www.sph.umn.edu/student-policies/. Students are expected to read and understand all policy information available at this link and are encouraged to make use of the resources available.

The University of Minnesota has official policies, including but not limited to the following:

- Grade definitions
- Scholastic dishonesty
- Makeup work for legitimate absences
- Student conduct code
- Sexual harassment, sexual assault, stalking, and relationship violence
- Equity, diversity, equal employment opportunity, and affirmative action
- Disability services
- Academic freedom and responsibility

Resources available for students include:

- Confidential mental health services
- Disability accommodations
- Housing and financial instability resources
- Technology help
- Academic support

EVALUATION & GRADING

Course evaluation will be based on class participation, homework assignments, a midterm exam, and a course project. The final grade is based on a weighted average score of a student’s performance in class participation, homework assignments, and a final project, with weights of 10%, 40%, and 50% respectively. There are about 3 homework assignments. Each assignment involves applying and evaluating some statistical learning methods, and/or writing a reading report; the students may need to do some more theoretical or computational problems, and read and critique journal articles. For the final project, possible topics include a case study (i.e. analysis of a specific data set), an empirical or theoretical comparison of a few statistical learning methods, or development/implementation and evaluation of a new/existing method (e.g. not covered or emphasized in class), or do an extensive literature review/survey on a topic. Your final project topic may be discussed with and approved by the instructor in advance. A project proposal will be due by Week 9. The project may be undertaken individually or by a team of 2-3 students, with a strong preference for collaboration. In the final 2-3 weeks, a presentation on each project will be given by its team members. A ≤ 5-page final project report for a whole team, including Introduction (or Background), Methods, Results, and possibly Discussion sections, is to be submitted during the last class session (tentatively on April 29). Each student is required to write a short critique on each presentation (not given on the same day as one’s own) and submit it with the same deadline as that for the final project report. No late homework assignments and project reports will be accepted unless with some legitimate reasons (e.g. illness with appropriate documents) or with my approval in advance.

Grading Scale
The University uses plus and minus grading on a 4.000 cumulative grade point scale in accordance with the following, and you can expect the grade lines to be drawn as follows:
<table>
<thead>
<tr>
<th>% In Class</th>
<th>Grade</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>93 - 100%</td>
<td>A</td>
<td>4.000</td>
</tr>
<tr>
<td>90 - 92%</td>
<td>A-</td>
<td>3.667</td>
</tr>
<tr>
<td>87 - 89%</td>
<td>B+</td>
<td>3.333</td>
</tr>
<tr>
<td>83 - 86%</td>
<td>B</td>
<td>3.000</td>
</tr>
<tr>
<td>80 - 82%</td>
<td>B-</td>
<td>2.667</td>
</tr>
<tr>
<td>77 - 79%</td>
<td>C+</td>
<td>2.333</td>
</tr>
<tr>
<td>73 - 76%</td>
<td>C</td>
<td>2.000</td>
</tr>
<tr>
<td>70 - 72%</td>
<td>C-</td>
<td>1.667</td>
</tr>
<tr>
<td>67 - 69%</td>
<td>D+</td>
<td>1.333</td>
</tr>
<tr>
<td>63 - 66%</td>
<td>D</td>
<td>1.000</td>
</tr>
<tr>
<td>&lt; 62%</td>
<td>F</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- **A** = achievement that is outstanding relative to the level necessary to meet course requirements.
- **B** = achievement that is significantly above the level necessary to meet course requirements.
- **C** = achievement that meets the course requirements in every respect.
- **D** = achievement that is worthy of credit even though it fails to meet fully the course requirements.
- **F** = failure because work was either (1) completed but at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I (Incomplete).
- **S** = achievement that is satisfactory, which is equivalent to a C- or better
- **N** = achievement that is not satisfactory and signifies that the work was either 1) completed but at a level that is not worthy of credit, or 2) not completed and there was no agreement between the instructor and student that the student would receive an I (Incomplete).
<table>
<thead>
<tr>
<th>Evaluation/Grading Policy</th>
<th>Evaluation/Grading Policy Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scholastic Dishonesty, Plagiarism, Cheating, etc.</td>
<td>You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (As defined in the Student Conduct Code). For additional information, please see <a href="https://z.umn.edu/dishonesty">https://z.umn.edu/dishonesty</a>. The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: <a href="https://z.umn.edu/integrity">https://z.umn.edu/integrity</a>. If you have additional questions, please clarify with your instructor. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class—e.g., whether collaboration on assignments is permitted, requirements, and methods for citing sources if electronic aids are permitted or prohibited during an exam. Indiana University offers a clear description of plagiarism and an online quiz to check your understanding (<a href="https://plagiarism.iu.edu/certificationTests/">https://plagiarism.iu.edu/certificationTests/</a>).</td>
</tr>
<tr>
<td>Late Assignments</td>
<td></td>
</tr>
<tr>
<td>Attendance Requirements</td>
<td></td>
</tr>
<tr>
<td>Makeup Work for Legitimate Reasons</td>
<td>If you experience an extraordinary event that prevents you from completing coursework on time and you would like to make arrangements to make up your work, contact your instructor within 24 hours of the missed deadline if an event could not have been anticipated and at least 48 hours prior if it is anticipated. University policy recognizes that there are a variety of legitimate circumstances in which students will miss coursework and that accommodations for makeup work will be made. This policy applies to all course requirements, including any final examination. Students are responsible for planning their schedules to avoid excessive conflicts with course requirements. 1. Instructors may not penalize students for absence during the academic term due to the following unavoidable or legitimate circumstances: illness, physical or mental, of the student or a student's dependent; medical conditions related to pregnancy; participation in intercollegiate athletic events; subpoenas; jury duty; military service; bereavement, including travel related to bereavement; religious observances; participation in formal University system governance, including the University Senate, Student Senate, and Board of Regents meetings, by students selected as representatives to those bodies; and activities sponsored by the University if identified by the senior academic officer for the campus or the officer's designee as the basis for excused absences. 2. Voting in a regional, state, or national election is not an unavoidable or legitimate absence. 3. Instructors are expected to accommodate students who wish to participate in party caucuses, pursuant to the Board of Regents resolution (see December 2005 Board of Regents Minutes, p 147.) 4. For circumstances not listed in (1), the instructor has primary responsibility to decide on a case-by-case basis if an absence is due to unavoidable or legitimate circumstances and grant a request for makeup work. Because this course is entirely online and all materials are available to students from the first day of the term, we expect students to plan accordingly if travel or access to the internet will cause them to miss a deadline. Note that our deadlines are generally set for 11:55 p.m. CST, so traveling to a different time zone will require additional planning. Further, circumstances that qualify for making up missed work will be handled by the instructor on a case-by-case basis; they will always be considered but not always granted. For complete information, view the U of M's policy on Makeup Work for Legitimate Absences (<a href="https://policy.umn.edu/education/makeupwork">https://policy.umn.edu/education/makeupwork</a>).</td>
</tr>
<tr>
<td>Extra Credit</td>
<td></td>
</tr>
<tr>
<td>Competency</td>
<td>Learning Objectives</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Evidence-based Approaches to Public Health</td>
<td>Apply suitable quantitative methods to analyze public health data</td>
</tr>
<tr>
<td>Public Health &amp; Health Care Systems</td>
<td>Discuss structural bias, social inequities, and health inequities</td>
</tr>
<tr>
<td>Planning &amp; Management to Promote Health</td>
<td>NA</td>
</tr>
<tr>
<td>Policy in Public Health</td>
<td>NA</td>
</tr>
<tr>
<td>Leadership</td>
<td>Apply negotiation and mediation skills to create a vision, empower others, and foster collaboration</td>
</tr>
<tr>
<td>Communication</td>
<td>Effective scientific communication both in writing and through oral presentation</td>
</tr>
<tr>
<td>Interprofessional Practice</td>
<td>Perform effectively on interprofessional teams</td>
</tr>
</tbody>
</table>