

School of Public Health

Syllabus and Course Information



UNIVERSITY OF MINNESOTA
Driven to DiscoverSM

PubH 7475/8475/Stat 8933 Statistical Learning and Data Mining Spring 2018

Credits: 3
Meeting Days: M&W
Meeting Time: 9:45am—11:00pm
Meeting Place: G55 Peik Gymnasium Hall
Instructor: Dr. Wei Pan
Office Address: A428 Mayo
Office Phone: 612-626-2705
Fax: 612-626-0660
E-mail: weip@biostat.umn.edu
Office Hours: 11am—12pm M&W

Instructor: Dr. Xiaotong Shen
Office Address: 384 Ford Hall
Office Phone: 612-624-7098
Fax: 612-624-8868
E-mail: xshen@umn.edu
Office Hours: 11am—12pm M&W

TA: Ms. Yan Li
Office Address: A446 Mayo
Office Phone:
Fax:
E-mail: lixx810@umn.edu
Office Hours: 11:15am--12:15pm Tu, 9—10 F

TA: Mr. Jun Young Park

Office Address: A446 Mayo
Office Phone:
Fax:
E-mail: park1131@umn.edu
Office Hours: 1pm—2pm M, 11am—12pm Th

I. Course Description

The subject of this course is closely related to machine learning, data analytics, and Big Data. This course will introduce various statistical techniques for prediction (i.e. learning) from data, possibly Big Data. Topics to be covered include penalized/regularized regression, linear discriminant analysis, tree-structured classifiers, neural networks, support vector machines, classifier ensembles (such as bagging and boosting), unsupervised learning, graphical models, recommender systems, and Big Data. These techniques can be applied in many fields, such as in business and bioinformatics/computational biology.

Course home pages: <http://www.biostat.umn.edu/~weip/course/dm/s18/home.html>
<http://users.stat.umn.edu/~xshen/stat8933.htm>

II. Course Prerequisites

Statistics at the level of PubH 7405–7407 or equivalent (e.g. Stat 5303) or permission of instructor, and some programming background in using R (or another high level language such as FORTRAN, C/C++, SAS).

III. Course Goals and Objectives

After taking the course, the student should have a working knowledge of using various machine learning techniques in practice.

IV. Methods of Instruction and Work Expectations

In-class lectures are the main method of instruction. Students are expected to come to class and participate in discussions, to do assignments, and to write a report respectively for the first and second parts and give a presentation for a course project towards the end of the semester. For those taking the 8000 section, some extra problems at a higher and theoretical level in the assignments and project are expected. Late assignments or project report are **not** accepted unless with advance permission from the instructor.

V. Course Text and Readings

Hastie T, Tibshirani R and Friedman J (2009). *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer, 2nd edition. Available on-line. (required)

James G, Witten D, Hastie T, Tibshirani R (2014). *An Introduction to Statistical Learning with Applications in R*. Springer. (not required; at a lower level with R examples)

VI. Course Outline/Weekly Schedule

- Dr. Pan:
 - Introduction (1 week);
 - Linear regression, penalized regression, logistic regression and discriminant analysis (2 weeks);
 - Classification and regression trees; support vector machines (1 week);
 - Ensemble methods, including Bagging, boosting, MART and RF (1 week);

- Neural networks (1 week);
- Unsupervised and semi-supervised learning: clustering, PCA, etc. (1.5 weeks);
- Dr. Shen: starting from week 9, March 19;
 - High-dimensional learning (1.5 weeks);
 - Graphical models and network analysis (1.5 weeks);
 - Recommender system (1 week);
 - Big data and unstructured data (1 weeks);
- Student presentations (2 weeks).

VII. Evaluation and 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 homework assignments, a mid-term (take-home) exam and a final project, with weights 40%, 20% and 40% respectively.**

There are about **5 homework assignments**. Each assignment involves applying and evaluating some statistical learning methods, and/or writing a reading report; those taking the 8000 section may need to do some more theoretical problems, and read and critique journal articles. We will use R programming language, which is freely available from <http://www.r-project.org>; you can use whatever language/system you like, though of course it will be your sole responsibility with programming. The mid-term exam is *tentatively* scheduled in week 8 (March 5, M). 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 an instructor in advance. The project will be done by a team of **three-four** students; *team work is strongly encouraged*. In the final one or two weeks, a short 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 due by 4pm (tentatively on May 11 Friday) in Dr. Shen's office or Stat main office. Each student is required to write a short critique on each presentation and submit 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.

A 4.000 - Represents achievement that is outstanding relative to the level necessary to meet course requirements

A- 3.667

B+ 3.333

B 3.000 - Represents achievement that is significantly above the level necessary to meet course requirements

B- 2.667

C+ 2.333

C 2.000 - Represents achievement that meets the course requirements in every respect

C- 1.667

D+ 1.333

D 1.000 - Represents achievement that is worthy of credit even though it fails to meet fully the course requirements

S Represents achievement that is satisfactory, which is equivalent to a C- or better.

For additional information, please refer to:

<http://policy.umn.edu/Policies/Education/Education/GRADINGTRANSCRIPTS.html>.

Course Evaluation

The SPH will collect student course evaluations electronically using a software system called CoursEval: www.sph.umn.edu/courseeval. The system will send email notifications to students when they can access and complete their course evaluations. Students who complete their course evaluations promptly will be able to access their final grades just as soon as the faculty member renders the grade in SPHGrades: www.sph.umn.edu/grades. All students will have access to their final grades through OneStop two weeks after the last day of the semester regardless of whether they completed their course evaluation or not. Student feedback on course content and faculty teaching skills are an important means for improving our work. Please take the time to complete a course evaluation for each of the courses for which you are registered.

Incomplete Contracts

A grade of incomplete "I" shall be assigned at the discretion of the instructor when, due to extraordinary circumstances (e.g., documented illness or hospitalization, death in family, etc.), the student was prevented from completing the work of the course on time. The assignment of an "I" requires that a contract be initiated and completed by the student before the last official day of class, and signed by both the student and instructor. If an incomplete is deemed appropriate by the instructor, the student in consultation with the instructor, will specify the time and manner in which the student will complete course requirements. Extension for completion of the work will not exceed one year (or earlier if designated by the student's college). For more information and to initiate an incomplete contract, students should go to SPHGrades at: www.sph.umn.edu/grades.

University of Minnesota Uniform Grading and Transcript Policy - A link to the policy can be found at onestop.umn.edu.

VIII. Other Course Information and Policies

Grade Option Change (if applicable)

For full-semester courses, students may change their grade option, if applicable, through the second week of the semester. Grade option change deadlines for other terms (i.e. summer and half-semester courses) can be found at onestop.umn.edu.

Course Withdrawal

Students should refer to the Refund and Drop/Add Deadlines for the particular term at onestop.umn.edu for information and deadlines for withdrawing from a course. As a courtesy, students should notify their instructor and, if applicable, advisor of their intent to withdraw.

Students wishing to withdraw from a course after the noted final deadline for a particular term must contact the School of Public Health Office of Admissions and Student Resources at sph-ssc@umn.edu for further information.

Student Conduct Code

The University seeks an environment that promotes academic achievement and integrity, that is protective of free inquiry, and that serves the educational mission of the University. Similarly, the University seeks a community that is free from violence, threats, and intimidation; that is respectful of the rights, opportunities, and welfare of students, faculty, staff, and guests of the University; and that does not threaten the physical or mental health or safety of members of the University community.

As a student at the University you are expected adhere to Board of Regents Policy: *Student Conduct Code*. To review the Student Conduct Code, please see: http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf.

Note that the conduct code specifically addresses disruptive classroom conduct, which means "engaging in behavior that substantially or repeatedly interrupts either the instructor's ability to teach or student learning. The classroom extends to any setting where a student is engaged in work toward academic credit or satisfaction of program-based requirements or related activities."

Use of Personal Electronic Devices in the Classroom

Using personal electronic devices in the classroom setting can hinder instruction and learning, not only for the student using the device but also for other students in the class. To this end, the University establishes the right of each faculty member to determine if and how personal electronic devices are allowed to be used in the classroom. For complete information, please reference:

<http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

Scholastic Dishonesty

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. (Student Conduct Code:

http://regents.umn.edu/sites/default/files/policies/Student_Conduct_Code.pdf) If it is determined that a student has cheated, he or she may be given an "F" or an "N" for the course, and may face additional sanctions from the University. For additional information, please see:

<http://policy.umn.edu/Policies/Education/Education/INSTRUCTORRESP.html>.

The Office for Student Conduct and Academic Integrity has compiled a useful list of Frequently Asked Questions pertaining to scholastic dishonesty: <http://www1.umn.edu/oscai/integrity/student/index.html>. If you have additional questions, please clarify with your instructor for the course. 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.

Makeup Work for Legitimate Absences

Students will not be penalized for absence during the semester due to unavoidable or legitimate circumstances. Such circumstances include verified illness, participation in intercollegiate athletic events, subpoenas, jury duty, military service, bereavement, and religious observances. Such circumstances do not include voting in local, state, or national elections. For complete information, please see:

<http://policy.umn.edu/Policies/Education/Education/MAKEUPWORK.html>.

Appropriate Student Use of Class Notes and Course Materials

Taking notes is a means of recording information but more importantly of personally absorbing and integrating the educational experience. However, broadly disseminating class notes beyond the classroom community or accepting compensation for taking and distributing classroom notes undermines instructor interests in their intellectual work product while not substantially furthering instructor and student interests in effective learning. Such actions violate shared norms and standards of the academic community. For additional information, please see: <http://policy.umn.edu/Policies/Education/Education/STUDENTRESP.html>.

Sexual Harassment

"Sexual harassment" means unwelcome sexual advances, requests for sexual favors, and/or other verbal or physical conduct of a sexual nature. Such conduct has the purpose or effect of unreasonably interfering with an individual's work or academic performance or creating an intimidating, hostile, or offensive working or academic environment in any University activity or program. Such behavior is not acceptable in the University setting. For additional information, please consult Board of Regents Policy:

<http://regents.umn.edu/sites/default/files/policies/SexHarassment.pdf>

Equity, Diversity, Equal Opportunity, and Affirmative Action

The University will provide equal access to and opportunity in its programs and facilities, without regard to race, color, creed, religion, national origin, gender, age, marital status, disability, public assistance status, veteran status, sexual orientation, gender identity, or gender expression. For more information, please consult Board of Regents Policy: http://regents.umn.edu/sites/default/files/policies/Equity_Diversity_EO_AA.pdf.

Disability Accommodations

The University of Minnesota is committed to providing equitable access to learning opportunities for all students. Disability Services (DS) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations.

If you have, or think you may have, a disability (e.g., mental health, attentional, learning, chronic health, sensory, or physical), please contact DS at 612-626-1333 to arrange a confidential discussion regarding equitable access and reasonable accommodations.

If you are registered with DS and have a current letter requesting reasonable accommodations, please contact your instructor as early in the semester as possible to discuss how the accommodations will be applied in the course.

For more information, please see the DS website, <https://diversity.umn.edu/disability/>.

Mental Health and Stress Management

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. University of Minnesota services are available to assist you. You can learn more about the broad range of confidential mental health services available on campus via the Student Mental Health Website: <http://www.mentalhealth.umn.edu>.

The Office of Student Affairs at the University of Minnesota

The Office for Student Affairs provides services, programs, and facilities that advance student success, inspire students to make life-long positive contributions to society, promote an inclusive environment, and enrich the University of Minnesota community.

Units within the Office for Student Affairs include, the Aurora Center for Advocacy & Education, Boynton Health Service, Central Career Initiatives (CCE, CDes, CFANS), Leadership Education and Development –Undergraduate Programs (LEAD-UP), the Office for Fraternity and Sorority Life, the Office for Student Conduct and Academic Integrity, the Office for Student Engagement, the Parent Program, Recreational Sports, Student and Community Relations, the Student Conflict Resolution Center, the Student Parent HELP Center, Student Unions & Activities, University Counseling & Consulting Services, and University Student Legal Service.

For more information, please see the Office of Student Affairs at <http://www.osa.umn.edu/index.html>.

Academic Freedom and Responsibility: for courses that do not involve students in research

Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study and to reserve judgment about matters of opinion, but they are responsible for learning the content of any course of study for which they are enrolled.*

Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact the instructor, the Department Chair, your adviser, the associate dean of the college, or the Vice Provost for Faculty and Academic Affairs in the Office of the Provost.

OR:

Academic Freedom and Responsibility, for courses that involve students in research

Academic freedom is a cornerstone of the University. Within the scope and content of the course as defined by the instructor, it includes the freedom to discuss relevant matters in the classroom and conduct relevant research. Along with this freedom comes responsibility. Students are encouraged to develop the capacity for critical judgment and to engage in a sustained and independent search for truth. Students are free to take reasoned exception to the views offered in any course of study and to reserve judgment about matters of opinion, but they are responsible for learning the content of any course of study for which they are enrolled.*

When conducting research, pertinent institutional approvals must be obtained and the research must be consistent with University policies.

Reports of concerns about academic freedom are taken seriously, and there are individuals and offices available for help. Contact the instructor, the Department Chair, your adviser, the associate dean of the college, or the Vice Provost for Faculty and Academic Affairs in the Office of the Provost.

* Language adapted from the American Association of University Professors "Joint Statement on Rights and Freedoms of Students".