

High Dimensional Data Analysis

University of Iowa

BIOS:7240

Spring 2021

Credit: 3 s.h.

Instructor: Prof. Patrick Breheny
Office: N336 CPHB
Phone: 384-1584
e-mail: patrick-breheny@uiowa.edu
Office hours: 2:30 p.m - 3:30 p.m. Tuesday
10:00 a.m - 11:00 a.m. Friday
Or e-mail to set up an appointment

Lecture: 9:30 a.m - 10:50 a.m.
Monday & Wednesday
S106AB CPHB

Course description: Increasingly, the data collected in many fields is high-dimensional, in the sense that many characteristics, or features, are recorded for each observation. The collection of this kind of data is a relatively recent phenomenon, and it poses many challenges that traditional statistical methods are incapable of addressing. This course will cover the analysis of high-dimensional data, with an emphasis on the use of penalized regression models. The primary topics of the course include large-scale hypothesis testing and the estimation of false discovery rates, the concept of penalized likelihood and its connections with Bayesian statistics, the lasso, elastic net, and nonconvex penalties such as MCP and SCAD. In addition, we will cover approaches to inference for penalized regression models, models with “structured sparsity” such as the group lasso, fused lasso, and hierarchical penalties, and the extension of penalties to generalized linear models, time-to-event models, and semi-parametric models.

Objectives: My objectives with this course are:

- To broaden your mind in terms of thinking about penalized likelihood methods, how they can be useful in statistics, and why they are particularly useful for high-dimensional data
- To introduce useful statistical methods for high-dimensional data
- To familiarize you with important topics in high-dimensional data that you may wish to research
- To bring you up to speed concerning terminology and concepts in high-dimensional data analysis and penalized regression so that you can more easily read research articles in the field

Textbook: *High-Dimensional Regression Modeling* (in progress), BREHENY, P. and HUANG, J. CRC Press.

Prerequisite: One year of graduate-level mathematical statistics (e.g., STAT 5100 and 5101), a basic course in regression (e.g., BIOS 5720), and a working knowledge of R.

Course website: The course notes, assignments, data sets, and other relevant materials will be made available on the course web site:

<http://myweb.uiowa.edu/pbreheny/7240/s21>

Homework: I am dividing the course up into “topics”, where each topic will last for one to two weeks, depending on the amount of material. There will be one homework assignment per topic. Each assignment will be divided into three sections: (1) Mathematical concepts and derivations (2) Simulation studies (3) Analysis of real data.

All assignments may be resubmitted for partial credit for the points lost on the original submission. The final grade will then consist of a weighted average of the original submission and the resubmission, with the resubmission receiving 1/3 of the weight. So for example, if Adam Smith got a 51/60 on the original submission, but a 60/60 on the resubmission, his final grade for the assignment would be 54/60. There is no deadline for the resubmissions, nor is there a limit to the number of times you may resubmit an assignment.

The assignments comprise your entire grade.

Computing: The simulation study and real data analysis portion of each project will involve a computer. I assume that you are familiar with the basics of R programming; I will introduce and demonstrate helpful additional functions and code during class.

Corrections: Despite my best efforts, I am sure that both the book and the notes have mistakes. If you spot a mistake, I very much want you to let me know about it so that I can correct it. I will award 1 bonus point (to be added to your homework total) for pointing out a typographical error and three bonus points for an error in content. Corrections will be made to the online version of the book/notes. Once an error has been corrected online, no more bonus points for that mistake are available.

Attendance: Regular attendance in this course is expected. No direct penalty will be applied for missing lectures. However, assignments will be based on what we cover in lecture, so skipping lecture is likely to hurt your grade (and, of course, your understanding of the material). And it will hurt my feelings.

Electronic communication: I will occasionally send e-mails to the class (to the account listed for you in the campus directory), so please check that account regularly.

Course schedule: See <http://myweb.uiowa.edu/pbreheny/7240/s21/notes.html> for the schedule of topics.

I look forward to getting to know you, and I hope that we have a great semester together.

COVID-19 considerations

Lectures: This course is designed to be flexible with respect to lectures being delivered either in person (in CPHB) or online (over Zoom), depending on the severity of the COVID-19 pandemic and other considerations. My plan is to discuss the format of this class with the students and decide on a plan after hearing everyone's input. All lectures, both in-person and remote, will be recorded and posted to ICON. Additional remarks concerning remote lectures:

- Most of you presumably are familiar with Zoom meetings by now. If not, or if you have any concerns about using this technology, please let me know – it is important that you are able to attend both office hours and remote lectures without technological barriers.
- If unexpected problems arise that prevent you from attending a lecture, please feel free to schedule an appointment if you still have questions after looking through the notes and watching the recording.
- For remote lectures, please turn on your camera so that I can see my students. Lecturing to a blank void that you can't see or hear is depressing.

Q & A: Normally, I love it when students come up and ask questions after class. Unfortunately, doing so is not prudent during a pandemic as the University tries to minimize congregation near entrances and exits. We (this involves effort from both you and I) should do our best to leave the room not much later than the scheduled time of 10:50. However, I certainly do not want to stifle questions, so please either (a) ask these questions during class time, (b) bring them up in office hours, or (c) ask them over e-mail.

Social distancing and mask policy: University policy dictates that students:

- Wear a face covering (mask or face shield) in all University buildings at all times
- Maintain an appropriate social distance of 6 feet or more, if possible
- Stay home if you have been exposed to, are diagnosed with, or experience symptoms consistent with COVID-19

If a student refuses to adhere to these requirements, I unfortunately have to cancel class and file a report with the Office of Student Accountability.

Changes: This syllabus reflects current plans, but things may change as the nature of the pandemic changes. I will communicate any changes to you promptly via e-mail, and update the online version of the syllabus accordingly.

Academic misconduct: You may discuss the assignments and your solutions with other students, but your writeup must be your own. Specifically, any copying of mathematical solutions, or copy/paste operations on typed documents or code are strictly forbidden. Doing so will be considered plagiarism. The University of Iowa takes plagiarism very seriously, and has in place a number of rather severe academic sanctions: <http://dos.uiowa.edu/policies/academic-misconduct>.

PhD in Biostatistics Competencies:

- Demonstrate an increased level of knowledge and understanding of current statistical theory, methods, and practices in the health sciences.
- Design, manage data, analyze and interpret data from a variety of experimental and observational studies.
- Communicate research findings, including new statistical methods developed, effectively to various audiences in writing and through oral presentation.

Concerns: Students with suggestions or complaints should see me first, and if we cannot come to an agreement, I will direct you to the Departmental DEO, Prof. Joseph Cavanaugh, N332 CPHB, joe-cavanaugh@uiowa.edu. Students may also contact the Undergraduate Program Director (if appropriate) or the Associate Dean for Academic Affairs in the College of Public Health. Another resource for students is the Office of the University Ombuds-person. If a complaint cannot be resolved at the departmental and/or collegiate level, students may file a formal complaint utilizing the procedure specified in Section II, Chapter 29.7 of the Operations Manual: <http://opsmanual.uiowa.edu>.

Accommodations for students with disabilities: The University of Iowa is committed to providing an educational experience that is accessible to all students. If you have a diagnosed disability or any other condition that would impair your ability to complete the course requirements as stated above, please inform me as early in the semester as possible, preferably at least two weeks prior to the scheduled activity. Students needing accommodations must register with Student Disability Services, (SDS): <https://sds.studentlife.uiowa.edu/students/apply> to obtain a Letter of Accommodation (LOA). The office is located at 3015 Burge Hall (319.335.1462). Students with complaints about disability accommodation should notify Student Disability Services and may be counseled to file a complaint with the Office of Equal Opportunity and Diversity. Read more about Disability Resources at <https://diversity.uiowa.edu/disability-resources>.

Administrative home: This course is given by the College of Public Health. This means that class policies on matters such as requirements, grading, and sanctions for academic dishonesty are governed by the College of Public Health. Students wishing to add or drop this course after the official deadline must receive the approval of the Associate Dean for Academic Affairs in the College of Public Health. Details of the University policy of cross enrollments may be found at: <https://www.provost.uiowa.edu/sites/provost.uiowa.edu/files/crossenroll.pdf>.

Severe weather: In severe weather, class members should seek appropriate shelter immediately, leaving the classroom if necessary. The class will continue if possible when the

event is over. For more information on Hawk Alert and the siren warning system, visit <http://hawkalert.uiowa.edu>.

Sexual harassment/misconduct: The University of Iowa prohibits sexual misconduct, dating/domestic violence, and stalking in any form, including sexual assault or sexual harassment, and any form of nonconsensual sexual conduct. Students should be able to live, study, and work in an environment free from all forms of sexual misconduct, dating/domestic violence, and stalking. Incidents of sexual misconduct can be reported to the Office of Sexual Misconduct Response Coordinator (OSMRC) or to the Department of Public Safety (DPS). If you are uncertain if what you have experienced from a student is sexual misconduct, see Section IV, Chapter 2 of the operations manual. If you are uncertain if what you experienced from a faculty or staff member is sexual harassment, see Section II, Chapter 4 of the operations manual. Students impacted by a Title IX issue (sexual misconduct, dating/domestic violence, or stalking) may be eligible to request an academic accommodation. Contact the OSMRC for assistance, definitions, and the full University of Iowa policy. If you or someone you know experiences sexual assault, sexual harassment, dating/domestic violence, stalking, or any other behaviors prohibited under this policy, you are strongly encouraged to seek assistance and support. See below for link to confidential resources.

- OSMRC: <https://osmrc.uiowa.edu>
- DPS: <https://police.uiowa.edu>
- Operations manual: <https://opsmanual.uiowa.edu>
- Confidential support: <https://osmrc.uiowa.edu/victim-resources/confidential-support>

Mental health: Students are encouraged to be mindful of their mental health and seek help if they are feeling overwhelmed and/or incapable of meeting course expectations. For assistance with the class, students are encouraged to talk to me. For additional support and counseling, students are encouraged to contact The University Counseling Service (UCS) in 1950 University Capital Centre Suite or 3223 Westlawn South. Call 319.335.7294 to schedule appointments. Find out more about the UCS at <http://counseling.uiowa.edu>. After hours, call the Johnson County Crisis Line at 319.351.0140 or 911 if you are having a mental health emergency.

Nondiscrimination in the Classroom and the College: The University of Iowa prohibits discrimination and harassment against individuals on the basis of race, class, gender, sexual orientation, national origin, and other identity categories set forth in the University's Human Rights policy. For more information, contact the Office of Equal Opportunity and Diversity at diversity.uiowa.edu. Students are invited to optionally share their preferred names and pronouns with their instructors and classmates. Students are also able to state this information in the [MyUI system under Student Information](#).

Free Expression: The University of Iowa supports and upholds the First Amendment protection of freedom of speech and the principle of academic freedom in order to foster a learning environment where open inquiry and the vigorous debate of a diversity of ideas are encouraged. Students will not be penalized for the content or viewpoints of their speech as long as student expression in a class context is germane to the subject matter of the class and conveyed in an appropriate manner.