High Dimensional Data Analysis

University of Iowa
BIOS:7600
Spring 2016
Credit: 3 s.h.

Instructor:	Prof. Patrick Breheny
Office:	N336 CPHB
Phone:	384-1584
e-mail:	patrick-breheny@uiowa.edu
Office hours:	If my office door is open, feel free to knock; Otherwise, please e-mail to set up an appointment.

Lecture:	1:00 p.m - 2:20 p.m. Monday & Wednesday S025A 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 semiparametric 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/7600/s16`

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 the book in its current state has 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. 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/7600/s16/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.

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`.

Complaints: 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, N312 CPHB, `joe-cavanaugh@uiowa.edu`. Students may also contact the Associate Dean for Education and Student Affairs in the College of Public Health. Another resource for students is the Office of the University Ombudsperson. 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`.

Disabilities: If anyone has a disability requiring special accommodations, please let me know as soon as possible, so that these arrangements can be made. For more information, visit: `http://sds.studentlife.uiowa.edu`.

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 and Student Affairs in the College of Public Health. Details of the University policy of cross enrollments may be found at: `http://test.sitenow.uiowa.edu/provost/files/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`.

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.

Sexual Harassment: Sexual harassment subverts the mission of the University and threatens the well-being of students, faculty, and staff. All members of the UI community have a responsibility to uphold this mission and to contribute to a safe environment that enhances learning. Incidents of sexual harassment should be reported immediately. The policy in its entirety may be found at Section II, Chapter 4 of the Operations Manual: `http://opsmanual.uiowa.edu`.

If you or someone you know may be a victim of 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. Assistance is available 24 hours a day, 7 days a week, from:

- Rape Victim Advocacy Program (RVAP) – confidential, certified victim advocacy services, 319-335-6000
- Domestic Violence Intervention Program (DVIP) –confidential, certified victim advocacy services, 319-351-1043or 800-373-1043
- Emergency Department, University of Iowa Hospitals and Clinics – confidential medical services, 319-356-2233
- University of Iowa Department of Public Safety – law enforcement services, 319-335-5022, or 911 from any campus phone

During business hours, you may also seek assistance from the University of Iowa Office of the Sexual Misconduct Response Coordinator at 319-335-6200.