AMS 380 – Data Mining

Spring 2023; 2:00 PM - 3:20 PM; Tue & Thu

Instructor Information

Instructor	Email	Office Location & Hours
Hyunwook Koh, Ph.D.	hyunwook.koh@stonybrook.edu	B521; 3:00 PM - 4:30 PM; Mon & Wed (or by appointment)

Course Information

Course Description

Statistical learning refers to a set of tools for making sense of complex datasets. In recent years, we have seen a staggering increase in the scale and scope of data collection across virtually all areas of science and industry. As a result, statistical learning has become a critical toolkit for anyone who wishes to understand data - and as more and more of today's jobs involve data, this means that statistical learning is fast becoming a critical toolkit for everyone. This course provides a broad and less technical treatment of key topics in statistical learning (e.g., sparse methods for classification and regression, decision trees, boosting, support vector machines, deep learning). Each chapter includes an R lab. This course is appropriate for anyone who wishes to use contemporary tools for data analysis.

Teaching Mode

All classes will be held in-person (**B313**).

Course Materials

All the course materials (e.g., lecture notes, programming notes, assignments) can be found on **Brightspace**.

Textbook

"An Introduction to Statistical Learning" by Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, 2^{ed} edition, Springer, 2021 (Required) (freely available at https://www.statlearning.com/)

"The Elements of Statistical Learning" by Trevor Hastie, Robert Tibshirani, Jerome Friedman, 2^{ed} edition, Springer, 2017 (Optional) (freely available at https://hastie.su.domains/ElemStatLearn/)

"Data Mining with R: Learning with Case Studies" by Luis Torgo, 2^{ed} edition, Chapman & Hall, 2017 (Optional)

Pre-requisite

AMS 311 (or instructor's permission: contact Prof. Hyunwook Koh)

Academic Integrity

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. For more comprehensive information on academic integrity, including categories of academic dishonesty, please refer to the academic judiciary website at http://www.stonybrook.edu/uaa/academicjudiciary/

School Policy on Attendance

1. If a student has over 20% unexcused absences, the student's final course grade will be an F.
2. Students should report the reason of absence to the professor in advance, or immediately after the absence.
3. When a student excuses his/her absence, the student must provide documentation of the reason for the absence to the professor.
4. The professor of the course reserves the right to excuse absences.
5. The professor may excuse the absence if the submitted documentation fulfills the following conditions: extreme emergences, severe medical reasons with doctor's note, very important events.

Critical incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Judicial Affairs any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn.

Course Evaluations

Stony Brook University values student feedback in maintaining the high-quality education it provides and is committed to the course evaluation process, which includes a mid-semester assessment as well as an end-of-the-semester assessment, giving students a chance to provide information and feedback to an instructor which allows for development and improvement of courses. Please click the following link to access the course evaluation system: http://stonybrook.campuslabs.com/courseeval/

Grading

Final grade = *f*(Attendance [10%] + Homework [10%] + Midterm 1 [25%] + Midterm 2 [25] + Final [30%]); ABCDF grading; 3 credits

Total score	Final grade
94 - 100	A
90 - 93	A-
87 - 89	B+
84 - 86	B

81 - 83	B-
78 - 80	C+
75 - 77	C
72 - 74	C-
69 - 71	D+
66 - 68	D
61 - 65	D-
≤ 60	F

Tentative Course Schedule

No.	Date	Topic	Homework
1	Feb 28	Introduction	TBA
2	Mar 2	Statistical Learning	TBA
3	Mar 7	Statistical Learning	TBA
4	Mar 9	Linear Regression	TBA
5	Mar 14	Linear Regression	TBA
6	Mar 16	Linear Regression	TBA
7	Mar 21	Classification	TBA
8	Mar 23	Classification	TBA
9	Mar 28	Classification	TBA
10	Mar 30	Midterm 1	
11	Apr 4	Resampling Methods	TBA
12	Apr 6	Resampling Methods	TBA
13	Apr 11	Linear Model Selection and Regularization	TBA
14	Apr 13	Linear Model Selection and Regularization	TBA
15	Apr 18	Linear Model Selection and Regularization	TBA
16	Apr 20	Moving Beyond Linearity	TBA
17	Apr 25	Moving Beyond Linearity	TBA
18	Apr 27	Moving Beyond Linearity	TBA
19	May 2	Tree-based Methods	TBA
20	May 4	Tree-based Methods	TBA
21	May 9	Tree-based Methods	TBA
22	May 11	Midterm 2	

No.	Date	Topic	Homework
23	May 16	Support Vector Machines	TBA
24	May 18	Support Vector Machines	TBA
25	May 23	Support Vector Machines	TBA
26	May 25	Deep Learning	TBA
27	May 30	Deep Learning	TBA
28	Jun 1	Deep Learning	TBA
29	Jun 6	No Class (Memorial Day)	
30	Jun 8	No Class (Reading Day)	
31	Jun 15	Final	

Exam Schedule

Date	Subject
Mar 30	Midterm 1
May 11	Midterm 2
Jun 15	Final