SYLLABUS

Statistical Learning with Big Data

STAT 6115/DSBA 6115, Fall 2020

Instructor

Name: Dr. Jun Song
Email: Jun.Song@uncc.edu
Office Hours: Thr. 2:00 pm – 4:00 pm (virtual/appointment)

Course Information

Meeting: M 5:30 pm - 8:15 pm, Sep. 19, 2020 - Dec. 15, 2020. (via Zoom)
Website: CANVAS.
Time zone: All announcements including date/time are based on the eastern time.
Textbook:
Required: An Introduction to Statistical Learning with Applications in R, by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (Link).
Recommended: The Elements of Statistical Learning, by Trevor Hastie, Robert Tibshirani, and Jerome Friedman
Prerequisite: STAT 5110 or STAT 5123
Description: This course covers many statistical learning methods which are relevant and useful in a wide range of academic and non-academic disciplines. By introducing principal ideas in statistical learning, the course will help students to understand conceptual underpinnings of methods in data mining. The topics include statistical learning; classical linear methods for regression and classification; cross-validation; bootstrap; modern linear methods; non-linear methods; tree-based methods; support vector machines; unsupervised learning. At the end of the course, students should have a basic understanding of how all of these methods work and be able to apply them to real-world problems.
Software to use: R. See the Software section detail.

Evaluation

Attendance: See Attendance section.
Homework: 30%
Quizzes: 15%
Midterm Exam: 25%
Final Project: 30%

Grading

A 90.00 % –
B 80.00 % – 89.99 %
C 70.00 % – 79.99 %
D 60.00 % – 69.99 %
F – 59.99 %
Required Work

Attendance
- You are expected to be present on time for all class sessions. You are responsible for all announcements and supplements given in class. The camera must be turned on during the class meetings.
- Attendance will be checked automatically via Zoom log report. If the duration is below 80% of the class meeting, then it will be considered to be an absence.
- You are allowed 3 excused/unexcused absences. Each additional absence will lower your final score by 3%.
- Distraction of class is not allowed. In this case, you will be out of the virtual meeting room immediately and it will count as an absence.

Homework
- There will be homework assignments for each chapter.
- Homework problem list will be announced at least a week before the due dates and it will be posted on CANVAS.
- The homework submission due is at the beginning of the class. For example, if the homework due is on September 14, then you need to submit the homework before attending the class on September 14. No late homework will be accepted unless you are told to do.
- You are allowed to drop the lowest homework score.

Pop Quizzes
- Pop quizzes will be given occasionally in class without announcement beforehand.
- Each quiz will take 15–30 minutes to complete.
- There will be no make-up quiz unless you have a proper documentation or a note from Dean of Students Office.
- You are allowed to drop the lowest quiz score.

Midterm Exam
- There is one in-class three hour-long midterm exam (open-book). Your camera must be turned on during the exam.
- The exam includes theory, application, and some computing.
- Date: November 2.
- There will be no make-up exam unless you have a proper documentation or a note from Dean of Students Office.

Final Project
- We will have a final project which includes a project proposal, presentation, and the project report.
- Project proposal (5 points) due by November 16.
- Project presentation (10 points): December 7 and 14.
- Project report (15 points) due by December 21.
Statistical Software R

R is a free statistical software and very strong in statistical analysis. In this course, we mainly use R for computation. It is a script-based language and easy to learn on your own. Python, Matlab and C++ are optional choices but not recommended for the homework, project, and the exam. Any other languages such as SAS will NOT be accepted.

R Introduction R Introduction by R Core Team
R Practice Data Camp Introduction to R

Course Policies

Syllabus The standards and requirements set forth in this syllabus may be modified at any time by the course instructor. Notice of such changes will be sent by CANVAS notice.

Announcement You are required to attend classes and check CANVAS course page on a daily basis. You are responsible for any announcements in class. In case of absence, students should let the instructor know about the absence in advance 48 hours before the class.

Communication
- Topic-related questions (normal in-class questions) can be asked through the chat. It will be answered during the class meeting.
- When you have questions while you are studying by yourself, you can use Discussion page of the Canvas to ask questions. They will be covered in class.
- When you have questions including private concerns, you can email me. Email questions will be responded within 2 business days, at the latest.

Policies for online course

Virtual Meeting All the classes will be recorded and saved in the cloud so that you can replay. The video is available up to 30 days and then it will be removed automatically.

Office Hours The office hour-meeting is based on the appointment. Meeting will be held via Zoom. All the meetings will be recorded but videos will not be shared/posted. Check the Canvas page for the details.

Disability Service

UNC Charlotte is committed to access to education. If you have a disability and need academic accommodations, please provide a letter of accommodation from Disability Services early in the semester. For more information on accommodations, contact the Office of Disability Services at 704-687-0040 or visit their office in Fretwell 230.

Academic Integrity

All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online.