Search...

Overview

Welcome to ASRM 499/552

Image credit: **oneillinois**

Instructor

- Name: Dr. Zhiyu Quan
- E-mail: zquan@illinois.edu
  - For easy reference, please use [ASRM 499/552] as the starting of the email subject.
  - Check and reply emails on Fridays, Mondays, and Wednesdays evening.
- ~~Office: 222 Illini Hall~~
- Website: http://zquan.pages.math.illinois.edu/
- Office hours:
  - TBA
  - **All date and time referenced in this course are in Central Time (CDT).**
  - Please feel free to make appointments if you cannot make it during office hours.
  - The instructor reserves the right to modify the course content and class schedule during the semester.

Teaching Assistant and Grader

- Name: TBD

Contents

- E-mail: TBD
  - For easy reference, please use [ASRM 499/552] as the starting of the email subject.
  - Check and reply emails on every weekdays' evening.

Course Infomation

- Learning are completely conducted online.
- The learning materials are uploaded to the Illinois Moodle learning management system, with the course space name:
  - "ASRM 499/552 G1 SP21: Predictive Analytics (Quan, Z)" (TBA)
  - link (NetID with password login is required): https://learn.illinois.edu/course/view.php?id=53100 (TBA)
  - It is recommended that you log on to the course website and check your e-mail frequently for updates, news and announcements.
- Zoom meetings: 2 lecture and 1 office hour
  - lecture: **6:30 PM - 7:50 PM Tuesdays** and **6:30 PM - 7:50 PM Thursdays** (if there is a pre-recorded lecture video, we will treat this meeting as a Q&A session)
  - office hour: TBA
- Zoom
  - Please use your real name, with first and last names, and turn on your camera for attendance and proper video conferencing etiquette. To secure your privacy, feel free to use a virtual background.
- Learning Objectives
  - Data are an essential source for insurance companies and financial institutions to learn about their risks and business models. The actuarial profession has a long tradition of dealing with data. Hang on, aren't actuaries the data scientists of the financial and insurance world? A more emphatic statement is: Actuaries was the first generation of the data scientist, under the definition of data science https://en.wikipedia.org/wiki/Data_science.

    This course will give an overview of the different statistical learning methods and algorithms that can be employed to discover useful information from datasets, to explain how to build a predictive model, and to communicate the results in a scientific report. We will cover case studies in different fields in finance and insurance. For instance, auto/home insurance claims, stock and option price prediction, mortality trends, health data analytics, etc.

The intuitive examples and theoretical (optional for undergraduate students) components in this course will help actuarial science students develop the programming skills for the Exam PA: Predictive Analytics https://www.soa.org/education/exam-req/edu-exam-pa-detail/ offered by the Society of Actuaries (SOA). It would be a great opportunity for statistics major students to gain the finance and insurance domain knowledge. Students receiving the grade of A+ will be enthusiastically recommended to the Illinois Risk Lab in Fall 2021.

- Textbook
    - **Regression Modeling with Actuarial and Financial Applications**, Edward W. Frees, 2010, New York: Cambridge. ISBN: 978-0-521-13596-2.
    - **Actuarial Statistics with R: Theory and Case Studies**, Guojun Gan and Emiliano Valdez, ACTEX Publications, Winsted, CT, 2018. URL: https://www.actexmadriver.com/orderselection.aspx?id=453143268 **Highly recommended for undergraduate students**.
    - **An Introduction to Statistical Learning, with Applications in R**, James, Witten, Hastie, Tibshirani, 2013, New York: Springer. A PDF of the text can be downloaded https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf
- Recommended References
    - Statistical Learning
        - **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**, Trevor Hastie, Robert Tibshirani, Jerome Friedman, https://web.stanford.edu/~hastie/ElemStatLearn/
    - R reference
        - **R for Everyone**, 2nd ed. Lander, 2017, Boston: Addison-Wesley, ISBN 978-0-13-454692-6
        - **Data Visualization: A Practical Introduction**, Healy, 2018, Princeton University Press, ISBN 978-0-691-18162-2. This book may be available as web pages at http://socviz.co/.

Preliminaries

- Preliminary knowledge about R is required. I attached “Introduction to R Programming” at the end of the syllabus.
- Knowledge about LATEX or (R Markdown or Jupyter Notebook) is helpful to conduct scientific reports.

Learning Objectives

- Topic 1: Predictive Analytics Problems and Tools
    - Introduction to statistical learning
    - Predictive modeling problems
    - Problem definition
    - Data visualization

    - Data types and exploration
    - Data issues and cleaning
- Topic 2: Supervise Learning
    - Linear model
    - Generalized linear models
    - Regularization
    - Tree-based models
    - Validation measures
    - Hyperparameters tuning
    - Case studies and interpretation
    - Other topics
- Topic 3: Unsupervise Learning
    - K-means
    - Hierarchical clustering
    - Principal component analysis.
    - Validation measures and hyperparameters specification
    - Case studies and interpretation
    - Other topics

Course Assessment

- Assessments are completely conducted online, which include 4 components: individual projects, oral examinations, take home midterm examination, and bonus points.

- Individual Projects (**60%**):

    - There are three components: classification (20%), regression (20%), and clustering (20%).
    - Rubric: model performance (5% + 5% + 5%) and scientific report (15% + 15% + 15%)
        - Model performance should be better than bench mark model (for example, regression model has larger R^2 compare to bench mark, say 0.6)
        - Scientific report (refer to PA rubric)
            - Communication
                - Executive summary – clearly and concisely written summary that is appropriate for someone who reads nothing else
                - Problem statement – clearly defines the problem and its business context
                - Use of tables and graphs – clearly constructed, labeled, and referenced
                - Interpretation of model results – relates the results of the modeling process to the problem statement
                - Audience – sections tailored to the audience as

        - described in the project statement
        - Code – easy to follow, using intuitive variable names and sufficient comments
    - Data Exploration and Feature Selection
        - Description of the data – summary statistics and graphs with interpretation
        - Identification of issues and corrective steps – includes handling missing data and possible transformations
        - Selection of features for use in the model – includes creating new features through transformations, clustering, or principal component analysis as appropriate.
        - Code – successfully runs and produces output presented in the report
    - Model Selection and Construction
        - Selection and justification of model type – relates model choice to the business problem and the available data
        - Estimation of model parameters and hyperparameters, with explanation – calibrates the selected model, including selecting features from the list previously established
        - Validation of the selected model – documents that an appropriate validation method was used and provides an estimate of model accuracy using previously unseen data
        - Description of selected model – describes the model in appropriate terms for the stated audience
        - Code – successfully runs and produces output presented in the report
- Your scientific report must be submitted via Moodle (preferably in PDF format).
- Your predicted results must be submitted via Kaggle InClass competitions https://www.kaggle.com/c/about/inclass. (I will share more details in the first lecture, It will be fun, I said)
- You are strongly encouraged to use LaTeX or (R Markdown or Jupyter Notebook) for the scientific report.
- If you submit a scanned copy of the scientific report, it is your responsibility to make sure that the scanned images are legible to the TA for grading.
- Any queries of scientific report grading are forwarded to the grader.
- Unless an unexpected or a special circumstance happens, and a written consent is obtained from the instructor by email before the due date and time, any assignment submissions after the due date and time are not graded.

    - Discussion is strongly encouraged in the Kaggle discussion boards, but plagiarism is strictly prohibited.
    - You are encouraged to form study groups to learn the materials in class or discuss about the project. However, you are forbidden from sharing your solution to be copied by other students. All submitted results must be your own work. Students who were found copying or sharing their solution to be copied by others will automatically receive a zero.
- Take Home Midterm Examination (**25%**)
- Oral Examination (**15%**):
    - Short interview/oral examination will be conducted via Zoom privately before the final week.
    - The exact oral examination date will be annouced week 12.
    - Oral examination is 15 min virtual meeting.
    - Instructor will pose questions to the student in spoken form and the student has to answer the question in such a way as to demonstrate sufficient knowledge of the subject to pass the exam.
    - The grade is available right after the virtual meeting.
    - Any queries of examination grading are forwarded to the instructor.
    - Oral examination is a good way to prepare students for future job/academic interviews.
    - Oral examination helps reduce the risk of academic dishonesty.
    - Google: tips to prepare for an oral examination.
- Bonus:
    - **5%** The attendance for all lectures (**turn on your camera**) and tutorial videos are expected and checked.
    - **1%** Randomly spot-checking that students are continually engaged, for example, active discussion in discussion board and random quiz during the lecture.
    - **5%/model** Best model performance in Kaggle InClass competitions. Theoretically, you could have had up to 30% bonus points if you won all six model competitions. (It will be fun, you will say)
- Final Grade
    - The final score is calculated based on the formula:
        - 60% Individual Projects + 25% Take Home Midterm Examination + 15% Oral Examination + Bonus.
    - The final grade is then mapped by the following table.

Final Grade	Final Score
A−, A, A+	90 – 100
B−, B, B+	80 – 90
C−, C, C+	65 – 80
D−, D, D+	50 – 65

F	0 – 50

Academic Integrity Statement

The University of Illinois at Urbana-Champaign Student Code should also be considered as a part of this syllabus. Students should pay particular attention to Article 1, Part 4: Academic Integrity. Read the Code at the following URL: http://studentcode.illinois.edu/.

Academic dishonesty may result in a failing grade. Every student is expected to review and abide by the Academic Integrity Policy: http://studentcode.illinois.edu/. Ignorance is not an excuse for any academic dishonesty. It is your responsibility to read this policy to avoid any misunderstanding. Do not hesitate to ask the instructor if you are ever in doubt about what constitutes plagiarism, cheating, or any other breach of academic integrity.

Accommodations Statement

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabil- ities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES, you may visit 1207 S. Oak St., Champaign, call 333-4603, e-mail disability@illinois.edu or go to the DRES website.

Emergency Response Statement

Emergency response recommendations can be found at the following website: https:// police.illinois.edu/emergency-preparedness/. I encourage you to review this web- site and the campus building floor plans website within the first 10 days of class. https: //police.illinois.edu/emergency-preparedness/building-emergency-action-plans/

Family Educational Rights and Privacy Act (FERPA) Statement

Any student who has suppressed their directory information pursuant to Family Educational Rights and Privacy Act (FERPA) should self-identify to the instructor to ensure protection of the privacy of their attendance in this course. See https://registrar.illinois.edu/ academic-records/ferpa/ for more information on FERPA.

Sexual Misconduct Policy and Reporting Statement

The University of Illinois is committed to combating sexual misconduct. Faculty and staff members are required to report any instances of sexual misconduct to the Universitys Title IX and Disability Office. In turn, an individual with the Title IX and Disability Office will provide information about rights and options, including accommodations, support services, the campus disciplinary process, and law enforcement options.

A list of the designated University employees who, as counselors, confidential advisors, and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here: https://wecare.illinois.edu/resources/students/ #confidential. Other information about resources and reporting is available here: https: //wecare.illinois.edu/.

Last updated on Aug 24, 2020