



ASTR 502: Data Mining and Machine Learning in Astronomy

Spring 2020

Monday/Wednesday 3pm – 3:50pm, SO 208

Description of Course

This is a graduate level elective course aiming at providing the interface between astronomical data analysis problems and modern statistics methods. Modern astronomy and astrophysics is undergoing a revolution with dramatic increases in both the volume and complexity of astronomical data. The last decade saw the emergence of many terabyte-level sky surveys across the electromagnetic spectrum; the next decade, data volumes will enter the petabyte regime, with an ever strong time domain component. These new data sets represent quantum leaps in our abilities for new astronomical discoveries, but also present significant challenges to standard analysis tools normally employed in astronomy.

The goal of this course is to bridge the gap between modern large data surveys and the data analysis tools that have been provided in normal graduate courses. The course will start with a brief review of the modern statistics framework relevant to large scale data analysis, including probabilities and statistical distribution, classical and Bayesian statistical inferences. Then it will cover the main topics of the course: data mining and machine learning, including density estimation, clustering analysis, dimensionality reduction, regression and model fitting, classification and time series analysis. Another key component of the course is to introduce commonly used data mining and machine learning tools, in the context of Python-based packages, which will be used in solving data problems throughout the course.

The course will be given in a combination of instructor lectures, student-led seminars and guest lectures. After the first section of introductory material, students will lead discussion and demonstration of most of the topics, and guest lectures will introduce important current and future key big data projects in astronomy. The class will conclude with final projects on using data mining and machine learning tools of your own research data.

Instructor and Contact Information

Prof. Xiaohui Fan

Office: SO 340

Phone: 626-7558

email: fan@as.arizona.edu

office hour: Friday 10am – 11am, SO 340, or by appointment

Course Format and Teaching Methods

The course will be given in a combination of instructor lectures, student-led seminars and guest lectures. After the first section of introductory material, students will lead discussion and demonstration of most of the topics, and guest lectures will introduce important current and future key big data projects in astronomy. The class will conclude with final projects on using data mining and machine learning tools of your own research data.

Course Objectives and Expected Learning Outcomes

1. Exhibit an expert-level facility to engage with the principle findings, common applications, current problems, fundamental techniques, and underlying theory of the astronomy discipline.
2. Demonstrate advanced discipline skills and knowledge necessary to utilize the observational techniques, instrumentation, computational methods, and software applications used to investigate modern astrophysical phenomena and problems.
3. Develop expertise with communicating, translating and interpreting fundamental astronomical concepts and research results in oral and/or written formats.
4. Conduct independent research and/or gain mastery-level knowledge of a specific area of the discipline of astronomy.
5. Engage in the scholarly, ethical, and discipline specific practices of the field at a professional level.

Absence and Class Participation Policy

The UA's policy concerning Class Attendance, Participation, and Administrative Drops is available at: <http://catalog.arizona.edu/policy/class-attendance-participation-and-administrative-drop>

The UA policy regarding absences for any sincerely held religious belief, observance or practice will be accommodated where reasonable, <http://policy.arizona.edu/human-resources/religious-accommodation-policy>.

Absences pre-approved by the UA Dean of Students (or Dean Designee) will be honored. See: <https://deanofstudents.arizona.edu/absences>

Participating in the course and attending lectures and other course events are vital to the learning process. As such, attendance is required at all lectures and discussion section meetings. Students who miss class due to illness or emergency are required to bring documentation from their health-care provider or other relevant, professional third parties. Failure to submit third-party documentation will result in unexcused absences.

Course Communications

<http://sancerre.as.arizona.edu/~fan/Home/AST502.html>
<https://github.com/UA-ast502-2019/classnotebooks>

Required Texts or Readings

The main text book is: "Statistics, Data Mining, and Machine Learning in Astronomy", by Ivezić, Connolly, VanderPlas and Gray, 2019, Princeton University Press.

The codes and figures used in the book can be found at astroML website: <http://www.astroml.org/>

Final Project

Students will be divided into several groups to work on a data mining and machine learning final project and present it to the entire class.

Grading Scale and Policies

Class Attendance: 33.3%

Classroom Lecture: 33.3%

Final Project: 33.3%

Requests for incomplete (I) or withdrawal (W) must be made in accordance with University policies, which are available at <http://catalog.arizona.edu/policy/grades-and-grading-system#incomplete> and <http://catalog.arizona.edu/policy/grades-and-grading-system#Withdrawal> respectively.

Scheduled Topics/Activities

Week 1: Introduction

Week 2/3: Astronomical Surveys, Tools and Statistics Refresher

Week 4-16: Student lectures and guest lectures.

Week 17: Final Project Presentations.

Discussion Topics include:

1. maximum likelihood estimator
2. non-parametric estimators
3. Bayesian parameter estimation
4. Bayesian model selection
5. MCMC
6. density estimation (non-parametric) - KDE, nearest neighbor
7. density estimation (parametric)
8. clustering
9. Dimensionality: PCA
10. Dimensionality: manifold learning and ICA
11. Regression: linear models
12. Regression: nonlinear models
13. Classification: generative
14. Classification: SVM, trees and forests
15. time series: basic models
16. time series: periodic
17. time series: localized and stochastic
18. Deep learning

Guest Lectures from:

1. Zeljko Ivezic: LSST overview
2. Ann Zabludoff: LSST Science
3. Stephanie Juneau/Robert Nikutta: Data Lab
4. Tom Matheson: ANTARES
5. CK Chan: Data Challenges in EHT
6. TBC

Classroom Behavior Policy

To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Threatening Behavior Policy

The UA Threatening Behavior by Students Policy prohibits threats of physical harm to any member of the University community, including to oneself. See <http://policy.arizona.edu/education-and-student-affairs/threatening-behavior-students>.

Accessibility and Accommodations

Our goal in this classroom is that learning experiences be as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, please let me know immediately so that we can discuss options. You are also welcome to contact the Disability Resource Center (520-621-3268) to establish reasonable accommodations. For additional information on the Disability Resource Center and reasonable accommodations, please visit <http://drc.arizona.edu>.

If you have reasonable accommodations, please plan to meet with me by appointment or during office hours to discuss accommodations and how my course requirements and activities may impact your ability to fully participate.

Please be aware that the accessible table and chairs in this room should remain available for students who find that standard classroom seating is not usable.

Students with Disabilities

If you anticipate barriers related to the format or requirements of this course, please meet with me so that we can discuss ways to ensure your full participation in the course. If you determine that disability-related accommodations are necessary, please register with Disability Resources (621-3268; <https://drc.arizona.edu/>) and notify me of your eligibility for reasonable accommodations. We can then plan how best to coordinate your accommodations.

Code of Academic Integrity

Students are encouraged to share intellectual views and discuss freely the principles and applications of course materials. However, graded work/exercises must be the product of independent effort unless otherwise instructed. Students are expected to adhere to the UA Code of Academic Integrity as described in the UA General Catalog. See: <http://deanofstudents.arizona.edu/academic-integrity/students/academic-integrity>.

UA Nondiscrimination and Anti-harassment Policy

The University is committed to creating and maintaining an environment free of discrimination; see <http://policy.arizona.edu/human-resources/nondiscrimination-and-anti-harassment-policy>

Subject to Change Statement

Information contained in the course syllabus, other than the grade and absence policy, may be subject to change with advance notice, as deemed appropriate by the instructor.