ASTR 3890 - Selected Topics: Data Science for Large Astronomical Surveys (Spring 2022)

Instructor: Nina Hernitschek, PhD
Lecture: Monday 12:30-2:00 pm (90 min)
Room: Buttrick Hall room 206
Office Hours: Data Science Institute (Sony Building) – Mo. 9am-10am, Fr. 1pm-2pm
Contact: nina.hernitschek@vanderbilt.edu

Course Description
This course will take a hands-on approach to data science methods for processing data from astronomical surveys. We will focus primarily on time-series data and cover topics and techniques related to data processing and analysis. The development of coding skills and best practices are additional goals of this course. Tasks and assignments will be conducted in Python. Covered topics include:
Astrostatistical and computational techniques for processing time-series data from large astronomical surveys.
Introduction in the usage of astronomical catalos such as SDSS and Gaia to retrieve time-series data.
Probability theory, comparison of frequentist and Bayesian inference. Strategies for data exploration and visualization.
Approaches to parameter estimation, model selection (e.g. Markov chain Monte Carlo) and machine learning using Python packages such as Scikit-Learn.

Course Format
This course is planned to be fully in-person for Spring 2022. If necessary, we will switch to Zoom.

Grading
Participation credit will be assigned by submitting your completed copy of the lecture Jupyter notebook. The completed lecture notebook and homework assignment must be submitted by 11:00am Central Time the following Monday. Credit is given for making a reasonable attempt at all tasks in the Jupyter notebook.

Final Take-Home Exam:
After the last session of the class I will assign a take-home exam that includes a series of statistics, coding and data analysis tasks that assess the course material.

Grading metric:
class participation and collaboration: 25%
homework assignments: 45%
final take-home exam: 30%
total: 100%

A = 90-100%, A-=85-90%, B+=80-85%, B=75-80%, B-=70-75%, C+=65-70%,C=60-65%, C-=55-60%, D=50-55%, F = less than 50%. Average numerical grades will be rounded to the nearest whole number.

Textbook
Statistics, Data Mining and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data - Ž. Ivezić, A. J. Connolly, J. T. VanderPlas, A. Gray
The textbook is available from the Princeton University Press website as well as from Amazon. Used books can be found on Amazon from about $20. An older version (1st edition) will do.

https://press.princeton.edu/books/hardcover/9780691198309/statistics-data-mining-and-machine-learning-in-astronomy
https://www.amazon.com/Statistics-Mining-Machine-Learning-Astronomy/dp/0691151687.

Lecture Material
all material, including homework assignments, is available in the github repository:
`https://github.com/ninahernitschek/astr3890`

Required Preknowledge
A working knowledge of Python will be necessary. ASTR 3890 code should be written in Python 3. One lecture will consist of a basic Python introduction. If you are not proficient in Python or have never coded before, it will be very useful to also complete a Python tutorial.

Personal Issues
To ensure that concerns are properly addressed from the beginning, if you have a physical, learning, or psychological issue or disability and require accommodations, please let me know as soon as possible. You must register with, and provide documentation of your disability to Student Access Services.

Lecture Schedule
Jan 24 Lecture 1: Introduction & Introduction to Python
Jan 31 Lecture 2: Astronomical Survey Data
Feb 7 Lecture 3: Introduction To Probability & Statistics: I
Feb 14 Lecture 4: Introduction To Probability & Statistics: II
Feb 21 Lecture 5: Classical/Frequentist Statistical Inference
Feb 28 Lecture 6: Bayesian Statistical Inference: I
Mar 5 - Mar 13 *Spring Break*
Mar 14 Lecture 7: Bayesian Statistical Inference: II
Mar 21 Lecture 8: Time Series Analysis: I
Mar 28 Lecture 9: Time Series Analysis: II
Apr 4 Lecture 10: Data Mining & Machine Learning: Intro to Scikit-Learn
Apr 11 Lecture 11: Dimensionality Reduction, Density Estimation & Clustering
Apr 18 Lecture 12: Classification: Introduction, Supervised Classification
Apr 25 Lecture 13: Unsupervised Classification