A598a: Machine Learning in Astronomy

Andrew Connolly and Mario Juric
University of Washington, Spring Quarter 2024

Class and Office hours

Class will be held Tuesdays and Thursdays in PAB B305 from 1.30 pm to 2.50 pm.

Office hours are on a drop-in basis and by appointment.

Class materials

Website and repository: https://github.com/uw-astro/astr-598a-spr24

Class JupyterHub:

Textbook: Ivezic, Connolly, VanderPlas & Gray "Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data"

Goals and Objectives

This course will introduce graduate students to common statistical and machine learning methods used in astronomy and other physical sciences. While it will include theoretical and methodological backgrounds for the techniques, the focus will be on the *application* of machine learning to astrophysical problems. Practical data analyses will be done using python tools, such as astroML module (see www.astroML.org), scikit-learn, and tensor flow/pytorch and applied to astronomical datasets. While focused on astronomy, this course should be useful to all graduate students interested in data analysis in the physical sciences and engineering. The lectures will be aimed at graduate students and the main discussion topics will be based on selected topics from Chapters 6-10, in the reference textbook "Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data".

The goal of this course is to give you the tools necessary to understand and analyze rich datasets, such as those from the SDSS to the Rubin and the LSST.

Homework and Final Project:

Homework will be to complete the course material and exercises presented within the class here. There will typically be two exercises per week. Homework will focus on practical applications using Python, designed to exercise what we have learnt in the week prior to the homework being assigned.

In place of a final exam, we will have a final project (a piece of software, or an analysis) to build or improve on using the techniques and libraries we learn about in the course. This project will be team-based with ~3 students per team. Projects will be defined in class. Projects can continue from the project you started in the "Introduction to Astrostatistics and Data-Intensive Astronomy" class.

Prerequisites:

Students taking this class are expected to have a background in calculus and Python and a basic understanding of statistics (e.g., Bayes theory and its application, and maximum likelihood estimation). Taking ASTR 598A: "Introduction to Astrostatistics and Data-Intensive Astronomy" (offered in the Autumn quarter) is strongly encouraged.

Course Structure

Lecture 1: Mar 26	Introduction to the course and project overview
Lecture 2: Mar 28	Regression: Linear, Non-linear, Outliers, Errors on Variables
Lecture 3: Apr 2	Regression: Basis Function Regression, Regularization, Cross-Validation
Lecture 4: Apr 4	Dimensionality Reduction
Lecture 5: Apr 9	Gaussian Processes
Lecture 6: Apr 11	Density Estimation: K-means, Kernel Density Estimation, K-nearest neighbors
Lecture 7: Apr 16	Density Estimation: Gaussian Mixture Models, eXtreme Deconvolution
Lecture 8: Apr 18	Generative classification: Naïve Bayes, GMM, LDA, ROC Curves
Lecture 9: Apr 23	Discriminative Classification: Decision Trees, Random Forests
Lecture 10: Apr 25	Hierarchical Bayes
Lecture 11: Apr 30	Approximate Bayesian Computation
Lecture 12: May 2	Deep Learning: Neural Networks
Lecture 13: May 7	Deep Learning: Convolutional Neural Networks
Lecture 14: May 9	Deep Learning: Autoencoders and Dimensionality Reduction
Lecture 15: May 14	Deep Learning: Attention and Transformers
Lecture 16: May 16	Project time
Lecture 17: May 21	Project time
Lecture 18: May 23	Project time
Lecture 19: May 28	Project time
Lecture 20: May 30	Project Presentations

Note that these topics and timing may change dependent on the progression and interests of the class. Additional topics we may cover based on interest include: Variational Inference, Probabilistic Programming.