ASTR 598: AstroStatistics and Machine Learning

Andrew Connolly & Željko Ivezić

University of Washington, Winter Quarter 2018

Location and Time: Tuesdays and Thursdays 11:00-12:20, **Room: XXX**

Office Hours: Any time when our office doors are open;
After class, or Tue and Thu mornings by appt. are the best.

Grading ??: 10 homeworks, 8% each; final exam: 20%
key: >90%=A, >80%=B, >70%=C, >50%=D.

Class web site: https://github.com/XXX

Reference textbook:
Ivezić, Connolly, VanderPlas & Gray: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (Princeton University Press, 2014)
See http://press.princeton.edu/titles/10159.html

Learning Goals:

This course will introduce students to most common statistical and computer science methods used in astronomy and other physical sciences. It will combine theoretical background with examples of data analysis based on modern astronomical datasets. Practical data analysis will be done using python tools, with emphasis on astroML module (see www.astroML.org). While focused on astronomy, this course should be useful to all students interested in data analysis in physical sciences and engineering. The lectures will be aimed at undergraduate students and the main discussion topics will be based on Chapters 4 and 5, and selected topics from Chapters 6-10, from the reference textbook.

By taking this course, students will develop basic understanding of topics such as robust statistics, hypothesis testing, maximum likelihood analysis, Bayesian statistics, model parameter estimation, the goodness of fit and model selection, density estimation and clustering, unsupervised and supervised classification, dimensionality reduction, regression and time series analysis. Most of these topics will be applied in class homeworks to analysis of astronomical data.

Prerequisites: The students taking this class are required to have basic calculus and basic python skills, as well as basic scientific measurements and statistics skills at the level of a freshman lab.

Lecture format: New material will typically be covered during the first class in a week, while the second class in a week will be more focused on practical data analysis work.

Class Schedule:

- WEEK 0 (Thu, Jan 4): Introduction (syllabus, literature, astroML)
- WEEK 1 (starting Jan 8):
  - Tue:
  - Thu:
- WEEK 2 (starting Jan 15): Introduction to statistics (probability, distributions, robust statistics, Central Limit Theorem, hypothesis testing).
  - Tue (Željko):
  - Thu:
- WEEK 3 (starting Jan 22): Maximum likelihood and applications in astronomy.
  - Tue:
  - Thu:
- WEEK 4 (starting Jan 29): Bayesian statistics and introduction to MCMC.
  - Tue:
  - Thu:
- WEEK 5 (starting Feb 5): Model parameter estimation and model selection.
  - Tue (Željko):
  - Thu (Željko):
- WEEK 6 (starting Feb 12): Time series analysis.
  - Tue:
  - Thu:
- WEEK 7 (starting Feb 19): Big data in astronomy.
  - Tue (Mario?):
  - Thu (Andy?):
- WEEK 8 (starting Feb 26): Dimensionality reduction and regression.
  - Tue:
  - Thu:
- WEEK 9 (starting Mar 5): Density estimation and clustering.
  - Tue (Mario?):
  - Thu (Mario?):
- FINAL EXAM: Mar 13 (Tue, 11:00-12:20, XXX): cake and closed book final exam.

Homework

There will be bi-weekly homeworks, based on a term-long survey data analysis project. We will use GitHub and Jupyter notebooks for HW submission.