Sample Syllabus: Data Science with Astronomy

Course Description: Data Science is changing the world as we know it with novel tools allowing exploration of several Petabytes of real-world data. This course focuses on different data analysis and statistical methods with a focus on intriguing and pressing questions in Astronomy. The course will also introduce basic-level machine learning tools and deploy them on real-world publicly available datasets collected from various sky surveys. The course focuses on students to excel in data analysis tools to prepare them for their own career path if that is in academia or industry.

Prerequisite(s): Basic Python and Unix skills, Astro 101
Instructor: Vishal Gajjar **Course Website:** gajjarvishal.com
Credit Hours: 3

Textbooks:

Book1: *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, **Author:** John R. Taylor

Book2: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*
Authors: Željko Ivezić, Andrew J. Connolly, Jacob T. VanderPlas, and Alexander Gray

Course Objectives:

At the completion of this course, students will be able to:

1. Use Python for data interpretation and manipulation
2. Learn various statistical tools to get data insights
3. How to separate "bad" data from "good" data using various clustering tools
4. Regression analysis
5. Learn basics of Machine Learning
6. Gain hands on astronomy research experience through projects
7. Understand *knowledge discovery* methods utilised in real-world data mining applications

Tentative Schedule:

Week	Content
Week 1	• Introduction to NumPy, SciPy, Astropy and data visualisation • Reading: Appendix A of Book 2 • Assignment: Write an Astropy based code and upload it to Git (6 groups)
Week 2	• Statistics of Univariate data through Globular Cluster luminosity function • Reading: Chapter 3 of Book 2; Chapter 4 of Book 1 • Assignment: Download Globular cluster datasets and report on their various properties
Week 3	• Statistics of multivariate data through Shapley galaxy dataset • Reading: Chapter 3 of Book 2; Chapter 4 of Book 1 • Assignment: Download Shapley galaxy datasets and report on various properties using multidimensional clustering

Week 4	• Statistical Distributions I: Poisson distribution with repeating Fast Radio Burst events • Reading: Chapter 11 of Book 1 • Assignment: Use reported FRB arrival times and measure average separation between consecutive radio bursts.
Week 5	• Statistical Distributions II: Normal distribution and Central Limit Theorem • Reading: Chapter 5 of Book 1 • Assignment: Test Central Limit Theorem by generating and combining random variables from different distributions using NumPy and SciPy.
Week 6	• Data curating and outliers rejections in radio astronomy • Reading: Chapter 6 of Book 1 • Assignment: Identify Radio Frequency Interference in radio observations from Breakthrough Listen archival datasets
Week 7	• Mid-term exam
Week 8	• Regression Analysis I: Curve fitting with pulsar spectra • Reading: Chapter 8 of Book 2 • Assignment: Using the ATNF pulsar catalogue, find spectral indices of radio pulsars
Week 9	• Regression Analysis II: Error estimation and propagation • Reading: Chapter 3 of Book 1 • Assignment: Using the reported errors on pulsar fluxes, propagate that to measure error on the spectral indices.
Week 10	• Bayesian Statistical Inference and model selection • Reading: Chapter 5 of Book 1 • Assignment: From the radial velocity measurements of GJ667C, use Bayesian inferences to report on the planet properties.
Week 11	• Machine Learning I: Introduction • Reading: Chapter 6 of Book 2 • Assignment: K-Neighbours for Photometric Redshifts with AstroML
Week 12	• Machine Learning II: Introduction of various classifiers • Reading: Chapter 9 of Book 2 • Assignment: No assignments, prepare for final exam.
Week 13	• Final Exam

Grade Distribution:

Assignments	40%
Mid-term	30%
Final Exam	30%