CS 109B/Stat 121B/AC 209B/CSCI 109B: Final Project
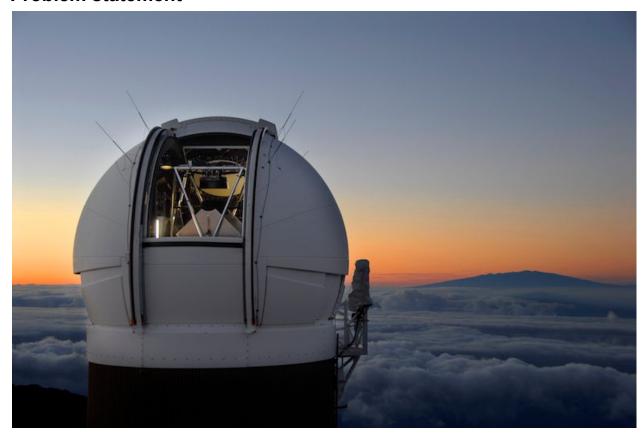Glickman, Protopapas

Module 1: A Real-Bogus Classifier for Astronomical Data

Module Instructors:
Matthew J. Holman, Senior Smithsonian Astrophysicist, Minor Planet Center Director
Matthew J. Payne, Minor Planet Center Project Scientist

Problem statement

Large data sets from ground-based and space-based observatories are transforming the way astronomical research is done. Examples include data from recent and ongoing surveys such as Pan-STARRS, ZTF, Kepler/K2, and TESS, as well as from future surveys such as LSST, SphereX, and NEOCam. Unlike other large, publicly available datasets you might have encountered, these are homogeneous, structured, and easily accessed and analyzed with community tools, largely in python. The scientists best positioned for research in this era are those who can adeptly bring simple, efficient tools and algorithms to bear across multiple, overlapping data sets to answer compelling scientific questions.

In this module, we will review some of the outstanding questions in astronomy and introduce some of the major data sets being used to answer those questions. We will then focus on the data from the Pan-STARRS survey and how those data can be used for a search for Planet-9.

Pan-STARRS-1 and Pan-STARRS-2 are a pair of 1.8m telescopes on Haleakala, Maui. Each telescope is equipped with a 1.4 gigapixel digital camera that can image 7 square degrees at a time. The images from these telescopes and camera generate petabytes of data each year. Unfortunately, the prototype detectors in these cameras generate a slew of false detections mixed in with detections of real, astrophysical sources. Your mission, should you choose to accept it, will be to develop a means of separating the real and bogus detections, i.e. classifying them. Doing so well will not only result in a cleaner data set but will support more rapid searches for slow-moving solar system objects.

Lectures

Lecture #1: A Golden Era for Astronomical Data
Lecture #2: Searching for Solar System Objects with Pan-STARRS-1 Data

Data Resources

We will provide a small data set to get started with the project, along with a set of python tools for accessing and manipulating the data.

References

The Pan-STARRS1 Surveys **https://arxiv.org/abs/1612.05560**

Using Machine Learning for Discovery in Synoptic Survey Imaging **https://arxiv.org/abs/1209.3775**

Machine Learning Based Real Bogus System for HSC-SSP Moving Object Detecting Pipeline, Lin et al. (2017) **https://arxiv.org/abs/1704.06413**