Exoplanet Identification from Light Curve Photometry

Overview/Problem Statement:

The discovery in 1992 of two planets orbiting a pulsar launched a large scale endeavor to search for, identify, and classify planets that do not orbit our sun. While there are different measurement methods for exoplanet searches, ultra-sensitive light photometry of stars have come to the fore in this endeavor. The technique involves recording the light from a star over an extended period of time (the light curve) and looking for periodic dips in the light. Periodic dips could potentially be caused by an exoplanet in regular orbit around the target star. The planet, while too small to resolve on its own, transits in front of the star and partially eclipses the star thus producing a tiny reduction in the light flux received by an observing telescope.

The problem is that planetary transits in a light curve are not always clean signals. There are various light curve noise sources (stellar variability from noise sources such as stellar spots or flares, instrument noise, etc) that can complicate true planetary transit identification. In addition, there are periodic false positives that can mimic true planetary transits. Among these are binary star systems where two stars orbiting each other can produce periodic dips in the light curve as one star transits in front of the other (or vice-versa). There are also stars whose luminosity varies periodically with time or light curves where the light measured from a star is periodically contaminated from a nearby object -- which could lead to false positives. We want to take light curve data, clean/filter as necessary, extract relevant features from these light curves, and construct a classifier that can distinguish between light curves with periodic transits that are planets vs. light curves with other kinds of light curves that are not of exo-planetary origin.

For training/testing our light curve classifier, we'll use light curve data and metadata from the Kepler Space Telescope Mission. This mission was active from 2009-2018. The Kepler space telescope observed > 500,000 stars and detected > 2500 planets.

Data Sources:

We'll be wrangling data from two sources:

1. NASA Exoplanet Archive API: Kepler Cumulative Table
   `'https://exoplanetarchive.ipac.caltech.edu/cgi-bin/nstedAPI/nph-nstedAPI?table=cumulative'`
   Contains Kepler star IDs, transit identification numbers, class labels, and useful transit parameters and metadata.
2. ExoMAST API: Light Curve Extraction: https://exo.mast.stsci.edu/docs/
   Contains the light curves which are downloaded by API requests.

Goals

1. Figure out what features in the light curve time series allow us to distinguish between planetary transits and false positives.
2. Extract these features, transform as necessary to make amenable to an ML classifier

3. Build transformation, hyperparameter optimization, and model selection pipeline to get best feature extraction/model performance for exoplanet/false-positive identification.