Goel, Amit

Statistics and Machine Learning based Outlier Detection Techniques for Exoplanets

Architectures of planetary systems are observable snapshots in time that can indicate formation and dynamic evolution of planets. The observable key parameters that we consider are planetary mass and orbital period. If planet masses are significantly less than their host star masses, then Keplerian Motion is defined as P^2 = a^3 where P is the orbital period in units of years and a is the orbital period in units of Astronomical Units (AU). Keplerian motion works on small scales such as the size of the Solar System but not on large scales such as the size of the Milky Way Galaxy. In this work, for confirmed exoplanets of known stellar mass, planetary mass, orbital period, and stellar age, we analyze Keplerian motion of systems based on stellar age to seek if Keplerian motion has an age dependency and to identify outliers. For detecting outliers, we apply several techniques based on statistical and machine learning methods such as probabilistic, linear, and proximity based models. In probabilistic and statistical models of outliers, the parameters of a closed form probability distributions are learned in order to detect the outliers. Linear models use regression analysis based techniques for detecting outliers. Proximity based models use distance based algorithms such as k-nearest neighbour, clustering algorithms such as k-means, or density based algorithms such as kernel density estimation. In this work, we will use unsupervised learning algorithms with only the proximity based models. In addition, we explore the relative strengths and weaknesses of the various techniques by validating the outliers. The validation criteria for the outliers is if the ratio of planetary mass to stellar mass is less than 0.001. In this work, we present our statistical analysis of the outliers thus detected.