Contribution ID: **31**   Type: **not specified**

Data Science in Relativistic Astrophysics - 1: Classification the stars using photometric optical data of SDSS

Thursday 4 November 2021 16:10 (45 minutes)

Classification the stars using photometric optical data of SDSS

RR Lyrae variables are periodic variable stars, commonly found in globular clusters. They are used as standard candles to measure (extra) galactic distances, assisting with the cosmic distance ladder. They are pulsating horizontal branch stars of spectral class A or F, with a mass of around half the Sun's. They are thought to have shed mass during the red-giant branch phase and were once stars of similar or slightly less mass than the Sun, around 0.8 solar masses. In contemporary astronomy, a period-luminosity relation makes them good standard candles for relatively nearby targets, especially within the Milky Way and Local Group. They are also frequent subjects in the studies of globular clusters. We use the set of photometric observations of RR Lyrae stars in the SDSS as our data. The data set comes from SDSS Stripe 82, and combines the Stripe 82 standard stars, which represent observations of non-variable stars; and the RR Lyrae variables pulled from the same observations as the standard stars, and selected based on their variability using supplemental data. The sample is further constrained to a smaller region of the overall color–color space following ($0.7 < u-g < 1.35$, $-0.15 < g - r < 0.4$, $-0.15 < r - i < 0.22$, and $-0.21 < i - z < 0.25$). These selection criteria lead to a sample of 92,658 non-variable stars, and 483 RR Lyraes. Two features of this combined data set make it a good candidate for testing classification algorithms:

1- The RR Lyrae stars and main sequence stars occupy a very similar region in u, g, r, i, z color space.

2- The extreme imbalance between the number of sources and the number of background objects is typical of real-world astronomical studies, where it is often desirable to select rare events out of a large background. Such unbalanced data aptly illustrates the strengths and weaknesses of various classification methods.

Our goal is to characterize the relation between the features in the data and their classes and apply these classifications to a larger set of unlabeled data. In this hands-on session, participants will learn how to use machine learning algorithms in practice and classify observed stars from optical data. This session has two parts in the first part we try to classify objects by some well known conventional machine learning algorithms such as logistic regression and etc. In the second part we use Neural Network for our classification purposes.

Presenter: Prof. ZHOLLIDEH HAGHIGHI, M. H. (IPM and KNTU, Iran)