ID de Contribution: **25**   Type: **Non spécifié**

Leveraging Robust Machine Learning Methods for Targeted Anomaly Searches

mardi 5 mars 2024 11:35 (25 minutes)

Anomalies are usually framed as "rare objects", lying in a low-density region of the feature space. However, finding them in practice under this broad definition can come with limitations: density estimation can be hard to perform reliably for high dimensional, noisy or complex (non-rectangular) data. Additionally, not all low-density points are interesting anomalies. In many cases, we are interested in a specific region of the feature space, looking for data points that diverge, in some precise aspects, from our expectations (e.g. if we have reliable models) or from otherwise similar data points. We can take advantage of robust, supervised ML methods for such anomaly searches without needing supervised anomalous examples. I will present results for the search of mid-infrared excess in FGK stars with a fully data-driven pipeline using Random Forests. To identify outlier candidates, we use a combination of the prediction errors and statistics using prediction errors of similar neighbouring points. This bypasses the need for accurate stellar models and fitting and provides a higher detection sensitivity, crucial in the mid-IR. This allows us to scale our search to an unprecedented data set of 4.9 million stars. Leveraging ML this way for targeted AD can be especially interesting in the absence of –good and cheap-to-compute– models to scale to large datasets. It is important to note that our approach detects outliers from the data perspective: if IR excesses were very common in our sample (e.g. young stars) and could be predicted from the input features, the model would learn this and our pipeline would not flag anomalies.

Orateur: Dr CONTARDO, Gabriella (SISSA)

Classification de Session: Contributed