What drives the scatter in the BPT diagrams? A Machine Learning based analysis

Mirko Curti*[1]

[1]Kavli Institute for Cosmology [Cambridge] – United Kingdom

Abstract

I will present a work aimed at identifying which physical properties are mostly connected with the position of local star forming galaxies in the classical diagnostic 'BPT' diagrams, by means of different Machine Learning algorithms.
Exploiting the huge statistics available from spectroscopic surveys of the local Universe like the SDSS, we have built a framework in which the dispersion of galaxies in the BPT diagrams and, in particular, their deviation from the best-fit curve of the star-forming locus, can be described by means of the relative variation in different observational properties compared to the average values retained by the bulk of the galaxies along the sequence. Artificial Neural Networks and Random Forest Trees are implemented to both classify whether galaxies lie above or below the sequence and to predict the exact distance/offset from the sequence itself. We achieve a high accuracy on the test sample in both classification and regression tasks, with no overfitting. Moreover, different approaches are followed to rank the parameters in terms of how much informative they are for the models. We show that parameters associated to the nitrogen-over-oxygen abundance ratio (N/O) are the most relevant in predicting the level of offset in the [N II]-BPT, whereas features related to the star-forming state of galaxies (like SFR) perform better in the [S II]-BPT. However, we also show how both the performances and the relative importance of each feature change as we consider different regions separately within the diagrams.
These models represent a valuable benchmark for high redshift galaxy samples in order to assess to what extent the physics that shape the local BPT diagrams is the same causing the offset seen in high-z sources or, instead, whether a different framework or even different physical mechanisms need to be involved.

*Speaker