SoC- Astronomical Data Modelling and Interpretation

Presentation

Aswin Suresh

July 2021

Table of Contents

Table of Contents

Overview

The following aspects of data science and their use in astronomy and astrophysics were explored:

- Use of **Python** in data science and basic manipulation of FITS files using **mean and median stacking.**
- Demonstrating the efficiency of algorithms (such as k-d trees) over usual code, in the context of **cross-matching** two catalogues containing different data pertaining to an object of interest, specifically super-massive black holes.
- **Basic SQL** commands (select, from where, group by, order by, having and so on) and joins (inner joins, outer joins, left and right joins). This was used to query a database containing data about exoplanets and their parent stars from the Kepler satellite.

Overview

- **Advanced SQL** commands: insert, delete, update. Created tables, primary and foreign keys. Combining python and SQL, and demonstrating ease of use of SQL over python for simple commands. This was also done in the context of exoplanets.
- Use of machine learning in data science and astrophysics. Demonstrating the power of a **regression classifier** (DTR from sci-kit learn) in calculating redshifts of galaxies. Training and evaluation sets of data and valuation metrics of a model (such as K-fold validation).
- Classification of galaxies using a **decision tree classifier**. Confusion matrices and random tree classifiers were introduced.

Table of Contents

Overview

The basics of astrophysics, starting from telescopes, to the solar system, ISM and planetary science, to stellar astrophysics, to high energy astrophysical systems/phenomena such as supernovae, neutron stars, black holes, and ending with galaxies and cosmology were covered. Here is a brief overview of what I learnt:

- History of astronomy, coordinate systems, telescopes and their types (UV, visible, X-Ray, IR, gamma-ray), observations from 21 cm line of Hydrogen and modern telescopes(adaptive optics and CCDs) and fluxes and magnitudes.
- Interstellar medium and its types (hot, warm and cold, their functions), protostellar and protoplanetary clouds, Jeans mass and planetary regions.

- Solar system and its planets, extrasolar planets and life, global warming, energy mechanisms in the Sun (thermonuclear synthesis, convection, radiation).
- H-R diagrams, open and globular star clusters and properties, death of stars and associated conditions, supernovae, formation of neutron stars and black holes and GRBs
- The milky way galaxy (structure $->$ globular clusters, halos and properties $->$ differential rotation), local groups, galaxies.
- Cosmology: The cosmological principle, friedmann equation, matter and radiation dominated models, redshifts,CMB, nucleosynthesis.
- Inflationary cosmology, dark matter, dark energy and concordance theories and lensing.

Table of Contents

Overview I

Read through a paper which described some machine learning tools used in astrophysics and data analysis in general. Topics covered include the following:

- What are input, validation and training datasets. The types of algorithms: supervised and unsupervised. Evaluation metrics: MAE, MSE, ROC, confusion matrix. Model parameters and hyperparameters (eg: best separation model in decision trees).
- Supervised algorithms: Use a predefined model
  - Support Vector Machine - separating classes in data using an N-dimensional hyperplane. Classification of nonlinear classes by transforming it into higher dimensions (kernel trick)

Overview II

- Decision trees and Random forests: Classifying data based on inequality conditions at every level (or node). However decision trees have a tendency to overfit data, and hence an ensemble is used which is called random forests. Probabilistic random forests also exist, which take care of data with high uncertainties.
- Artificial neural networks: Consists of a series of layers beginning with input and ending with output and a bunch of hidden layers in the middle. Applies linear transformations to the data through each layer, which is then acted upon by a non-linear activation function.

- Unsupervised algorithms: Discover new models. Typically used in astronomy for clustering (k-means), dimensionality reduction (PCA) and anomaly detection.