Modelling the Universe with Machine Learning

António da Silva

Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa

Statistical learning algorithms are at the heart of the so called 4th paradigm of scientific discovery. These are starting to emerge in Cosmology and Astrophysics as one of the most powerful tools to model and reveal the mysteries of the expanding Universe that is mostly made of unknown forms of dark energy and dark matter.

Extremely large datasets, such as galaxy surveys being collected by ground telescopes and the Euclid space mission[1], require intensive data analysis and new algorithms that act in a unsupervised way or without being explicitly instructed to do so. At the IA Cosmology group we propose the application of, state-of-the-art, machine learning techniques to address the following most pressing topics of modern cosmology.

Cosmological parameter inference

Traditional parameter inference typically uses Bayesian methods to estimate the "posterior" probability of parameters given an observed dataset (Fig. 1 solid lines) . These often involve extremely time consuming computations of a complex likelihood function over a multi-dimensional parameter space. New artificial neural network algorithms are able to provide almost instantaneous posterior parameter probabilities, without likelihood computations, (Fig. 1 yellow regions) after supervised training. Students are invited to apply this approach, using galaxy clusters as a probe of cosmology.

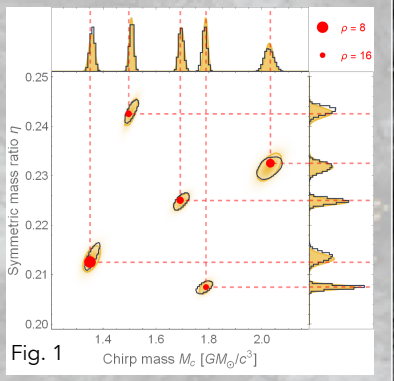

Fig. 1

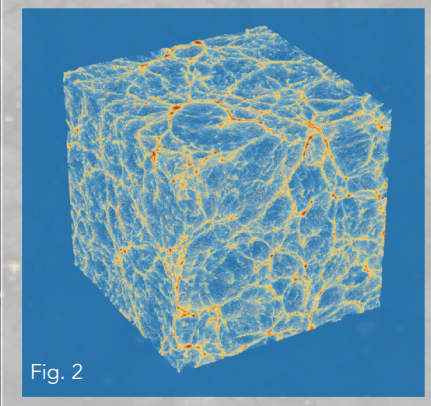

Fig. 2

Fast model regression and emulation

3D modeling of the Universe is now also possible with new supervised learning algorithms (Fig. 2). These can be combined with numerical N-Body simulations to provide fast and accurate representations of how galaxies, galaxy clusters, filaments and voids form and evolve in the 3D Universe. The students are invited to use these algorithms to produce regression models and fast emulators that allow the computation of key quantities such as the matter power spectrum and bispectrum for different dark energy / dark mater cosmology models.

Unsupervised galaxy cluster and void identification

Clusters and voids are peaks and valleys in the density distribution of the Universe (Fig.3). Unsupervised machine learning techniques are required to try to identify them without using *a priori* modeling or "manual" procedures that are impossible to apply to the vast datasets provided by galaxy surveys. Here, students are challenged to participate in the development and application of fast unsupervised algorithms for the detection of voids and clusters with real and simulated data.

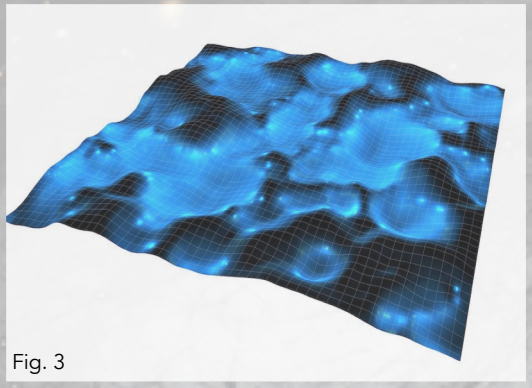

Fig. 3

Acknowledgments: AdS acknowledges FCT funding from contract Ref. IF/01135/2015 and IA multi-annual funding, contract ref. UIDB/04434/2020 and ref. UIDP/04434/2020;
Fig. 1 is from reference: Chua & Vallisneri 2019, doi: 10.1103/PhysRevLett.124.041102 ;

[1] Euclid Space Mission (ESA) with the participations of IA/FCUL: https://www.euclid-ec.org