GAUSSIAN PROCESSES AND NEURAL NETWORKS APPLIED TO PHOTOMETRIC REDSHIFT RECONSTRUCTION

Priscila Gutierres
IME - USP
priscila.gutierres@usp.br

Goals

- Show why machine learning algorithms have become much popular among various fields, including astronomy.
- We aim to perform an analysis of photometric redshifts estimated by using ANNz and GPz algorithms to estimate the photometric redshift in simulations and real data.

Photo-z redshift

Looking at higher redshifts is equivalent to looking back in time: they improve the studies of cosmology, expanding our knowledge of the universe.

Shortly, a spectrum is a graph of the amount of how bright the object is at different wavelengths.
It allows us to study various phenomena, i.e., test the principle of equivalence, because the laws of physics are the same everywhere.
All methods used for photo-z estimation require a spectroscopic data set for training and/or calibration.

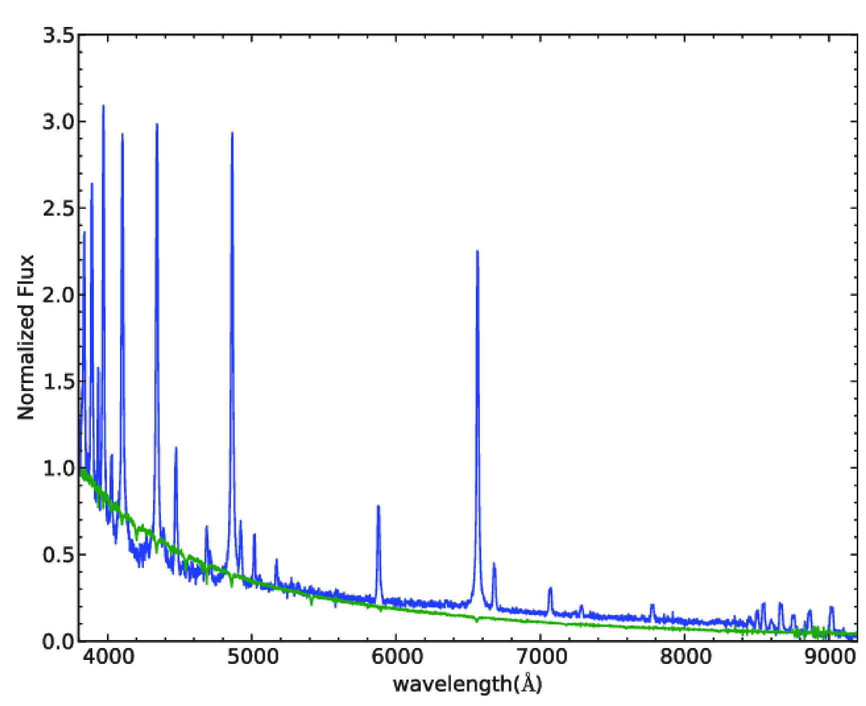

Figure 1: Spectral comparison between a CV star and a normal star. [5]

Dealing with astronomical data sets

Astronomical data sets are undergoing a rapid growth in size and complexity as past, ongoing and future surveys produce massive multi-temporal and multi-wavelength data sets, with a huge information to be extracted and analysed.
Today, galaxy large scale structure surveys rely on spectroscopic redshifts to produce high precision power spectrum measurements of the galaxy distribution.
The alternative to a full spectroscopic survey is to obtain multi-colour images of the sky and perform photometric redshift estimates for the galaxies we have available. [1]
When dealing with this problem, there are two main approaches: model-driven data analysis (template fitting methods) and data-driven analysis, which can use machine learning methods.
We will use data-driven analysis, more specifically Gaussian processes, **GPz**, and neural networks, **ANNz2**.

ANNz2

ANNz2 [4] is an updated version of the original **ANNz**, which uses artificial neural networks (ANNs) to estimate the photometric redshifts of galaxies.
This algorithm uses artificial neural networks and boosted decision/regression trees.
The TVMA method used by ANNz2's neural network is a multilayer perceptron, organized into at least three layers: the *input layer*, a a hidden layer and the *output layer*.

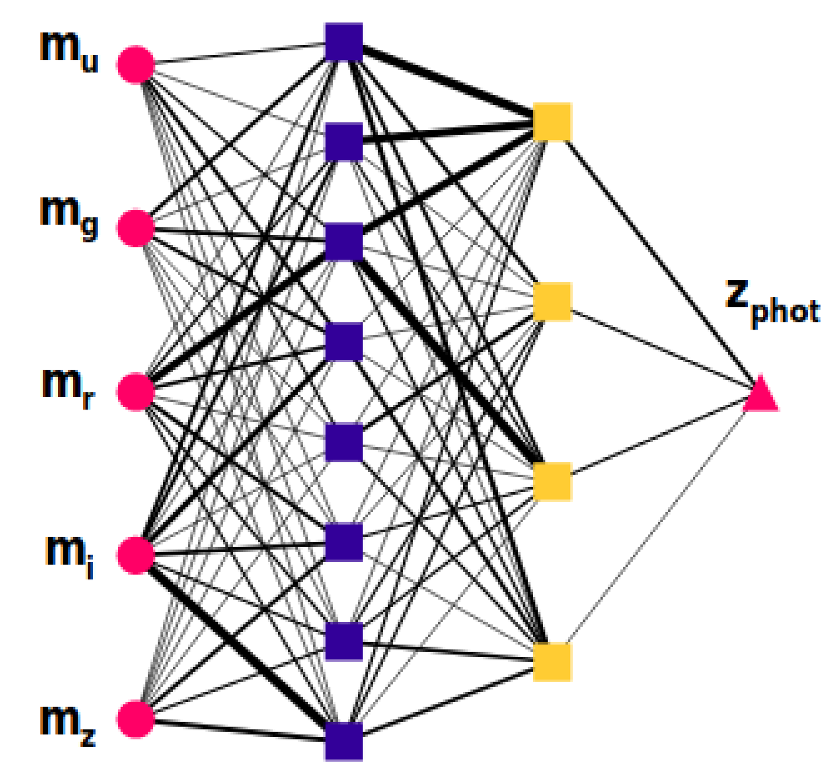

Figure 2: Schematic representation of an artificial neural network [4]

Also, there are boosted decision trees (BDTs) and other methods, line k-nearest neighbours and SVM.
ANNz2 utilizes multiple machine learning methods to properly derive the associated uncertainties and to produce both single-value solutions and PDFs.
The code is available for download at https://github.com/IftachSadeh/ANNz.

GPz

GPz [2] is a machine learning approach which uses sparce Gaussian processes (GPs) to estimate a photometric redshift and its variance. The GPs are probabilistic models for regression. The assumption underlying the GP approach is that there existis a function, f, to map a set of target inputs X (i.e. the galaxy photometry) ont a set of target outputs Y (i.e. the galaxy redshifts), such that $Y = f(X) + \epsilon$) [3]. Classical Gaussian processes (GPs) assume that the noise uncertainty is constant and only model the uncertainty about the mean function.
In GPz approach, the variance is an input-dependent function and is learned together with the mean function.
In a related work, described by Almosallam et al. 2016, sparse Gaussian processes for photometric redshift inference were proposed.

Conclusion

Using photometric redshifts we can probe much larger volumes of cosmological data than using spectroscopic ones, but doing this we get large measurement uncertainties.
Many machine learning methods have been used to estimate them, but we still need spectroscopic data to train our methods, which probe a limited volume of data compared to photometric ones.
In this study, it has been used GPz and ANNz2, which are used by various researchs to analyse a bunch of wide known surveys.

References

[1] **Abdalla, F. B. et al** *A comparison of six photometric redshift methods applied to 1.5 million luminous red galaxies*. (2011).

[2] **Almossallam, Ibrahim A. et al** *Using Sparse Gaussian Processes for Predicting Robust Inertial Confinement Fusion Implosion Yields*. (2016)

[3] **Rivera, J. D. et al**. *Degradation analysis in the estimation of photometric redshifts from non-representative training sets*. (2018).

[4] **Sadeh, Iftach et al** *ANNz2 - photometric redshift and probability distribution function estimation using machine learning* (2015)

[5] **Wei, Peng Luo, A-Li Li, Yinbi Pan, Jingchang Tu, Liangping Jiang, Bin Kong, Xiao Shi, Zhixin yi, Zhenping Wang, Fengfei Liu, Jie Zhao, Yongheng**. *Mining unusual and rare stellar spectra from large spectroscopic survey data sets using the outlier-detection method* (2013).