

Kernel PCA for Supernovae Photometric Classification

Emille E. O. Ishida

IAG, Universidade de São Paulo, Rua do Matão 1226, Cidade Universitária, CEP 05508-900,
São Paulo, SP, Brazil
email: emilleishida@usp.br

Abstract. In this work, we propose the use of Kernel Principal Component Analysis (KPCA) combined with $k = 1$ nearest neighbor algorithm (1NN) as a framework for supernovae (SNe) photometric classification. It is specially recommended for analysis where the user is interested in high purity in the final SNe Ia sample. Our method provide good purity results in all data sample analyzed, when $\text{SNR} \geq 5$. As a consequence, we can state that if a sample as the *Supernova Photometric Classification Challenge* were available today, we would be able to classify $\approx 15\%$ of the initial data set with purity higher than 90%. This makes our algorithm ideal for a first approach to an unlabeled data set or to be used as a complement in increasing the training sample for other algorithms. Results are sensitive to the information contained in each light curve, as a consequence, higher quality data (low noise) leads to higher successful classification rates.

Keywords. (stars:) supernovae: general, methods: statistical, techniques: photometric

1. Introduction

Type Ia supernovae (SNe Ia) observations provided the first evidence towards an accelerated cosmic expansion in late 1990's. Following this first result the scientific community is deeply engaged in improving these cosmological results in order to have a better idea about the properties and nature of dark energy. Such line of thought has lead to implementation of a large number of on going and future SNe surveys, aimed at observing SNe Ia for cosmological use. The result of such effort can already be experienced. According to Sako *et al.* (2011), within 3 observational seasons, the *Sloan Digital Sky Survey* (SDSS) supernova survey was able to obtain ≈ 3500 SNe light curves. However, given the limited spectroscopic follow-up available, for only ≈ 500 of them the spectroscopic confirmation was obtained. At this point, only this small sub-set is used in cosmological analysis. This situation will become more drastic with the advent of surveys now under construction.

In order to put different SNe classifiers in the same grounds, a *Supernova Photometric Classification Challenge* (SNPCC) was proposed. It consisted of sample of synthetic light curves, build according to what is expected from the 5 years of operations of the *Dark Energy Survey* (DES). The data was divided into a spectroscopically confirmed sample (≈ 1000), for which the labels were provided, and a photometric only sample (≈ 20000). Results were reported in Kessler *et al.* (2010).

2. Kernel PCA application

In this work, we approach the SNe photometric classification problem using a 2 steps machine learning algorithm. In a first moment, we applied *Kernel Principal Component Analysis* (KPCA) to the light curves in the spectroscopic sample. KPCA is a statistical

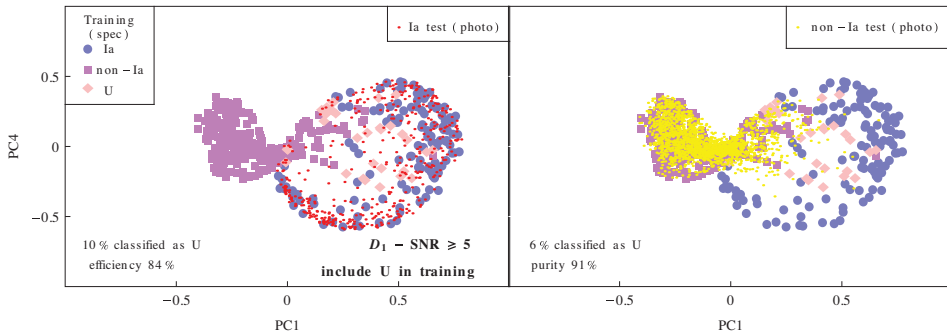


Figure 1. Example of low dimensional representation. $D_1 + \text{SNR} \geq 5$ means we used data between -3 and +24 days since maximum brightness demanding $\text{SNR} \geq 5$ for at least epochs in each filter. The pink triangles correspond to SNe in the boundaries of Ia/nonIa regions.

tool aimed at reducing the dimensionality of a given parameter space. In our suggested framework, all available information for a given SNe (flux measurements in different epochs and filters) are mapped into a 2-dimensional parameter space where the *locus* occupied by Ia are as separate as possible to that occupied by nonIa. Once this lower dimensional representation is constructed, we take one photometric light curve at a time and project it in this new parameter space. A $k=1$ Nearest Neighbor algorithm is then used to address a classification to this photometric point based on the geometrical distribution of the spectroscopic sample (see example of point projections in figure 1).

The method was applied to the SNPCC data set. Although our results are sensitive to low SNR observations, leading to a non-ideal efficiency at high redshifts, it achieved impressive purity values when compared to results from other algorithms reported within the *Challenge*, specially in redshifts within $0.2 \leq z < 0.5$. The good purity values are important given that, for a cosmological analysis, nonIa contamination in a Ia sample can lead to wrong cosmological models. A detailed analysis of these results and comparisons to others already available in the literature can be found in Ishida & de Souza (2012).

3. Conclusions

We proposed the use of KPCA+1NN algorithm as a framework for SNe photometric classification. From applying our method to the SNPCC data set, we are able to state that, if a sample as the one expected from the DES were available today, we would be able to classified 15% of the original set with purity levels $\geq 90\%$. This makes our method ideal as a first approach to the already available light curves samples without spectroscopic confirmation, as the one within the SDSS data.

Acknowledgements

E.E.O.I. is partially supported by the Brazilian financial agency FAPESP through grant number 2011/09525-3.

References

- Ishida, E. E. O. & de Souza, R. S. 2012, arXiv:astro-ph/1201.6676
 Kessler, R., *et al.* 2010, *PASP* 122, 898
 Sako, M., *et al.* 2011, *ApJ* 738, 162