Kernel PCA for SNe photometric classification

Emille E. O. Ishida
IAG – University of Sao Paulo, Brazil

In collaboration with Rafael S. De Souza (KASI)

astro-ph/1201.6676

IAU General Assembly, August/2012, Beijing, China

1. The problem

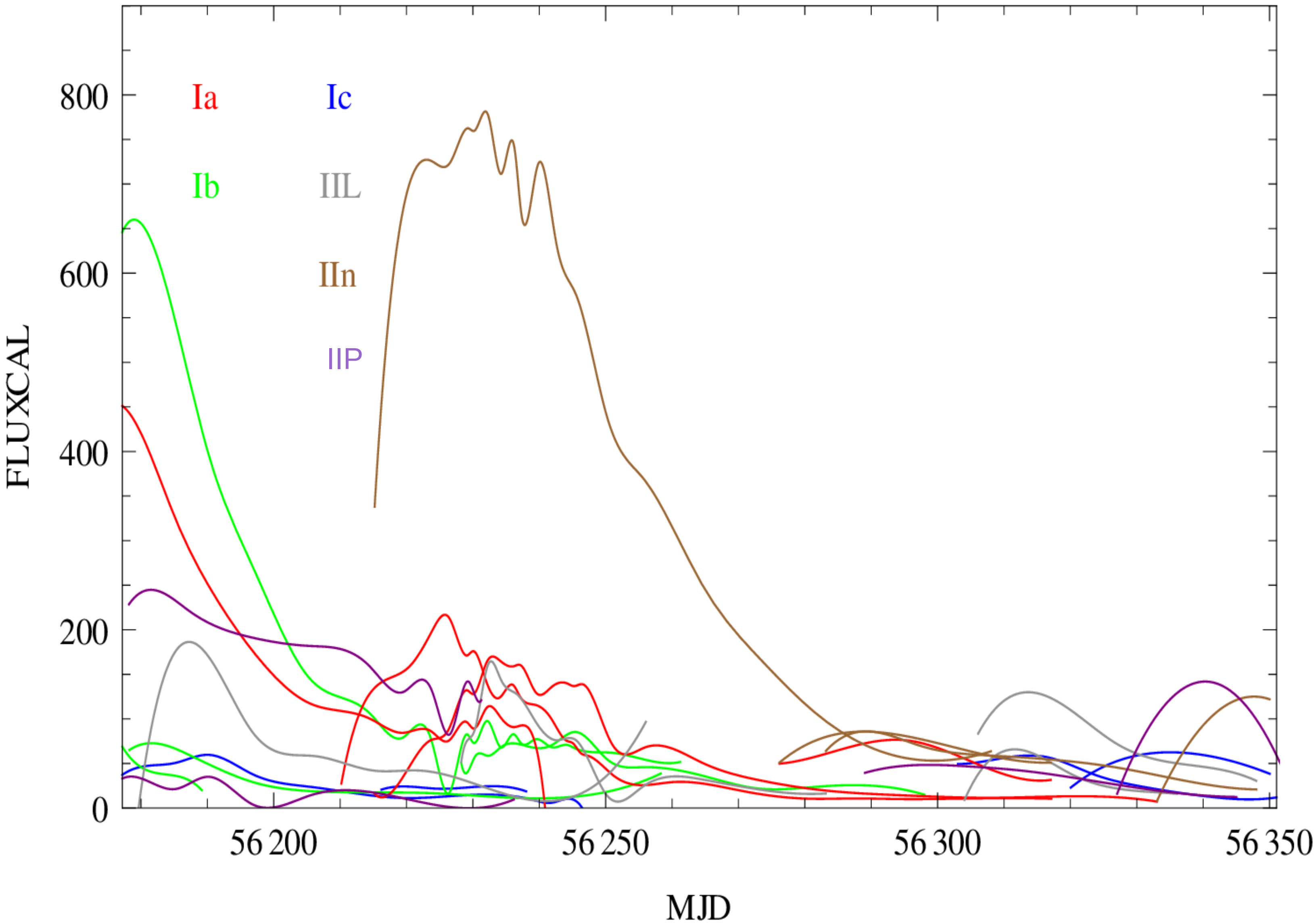
Ironically...

We have too much data!

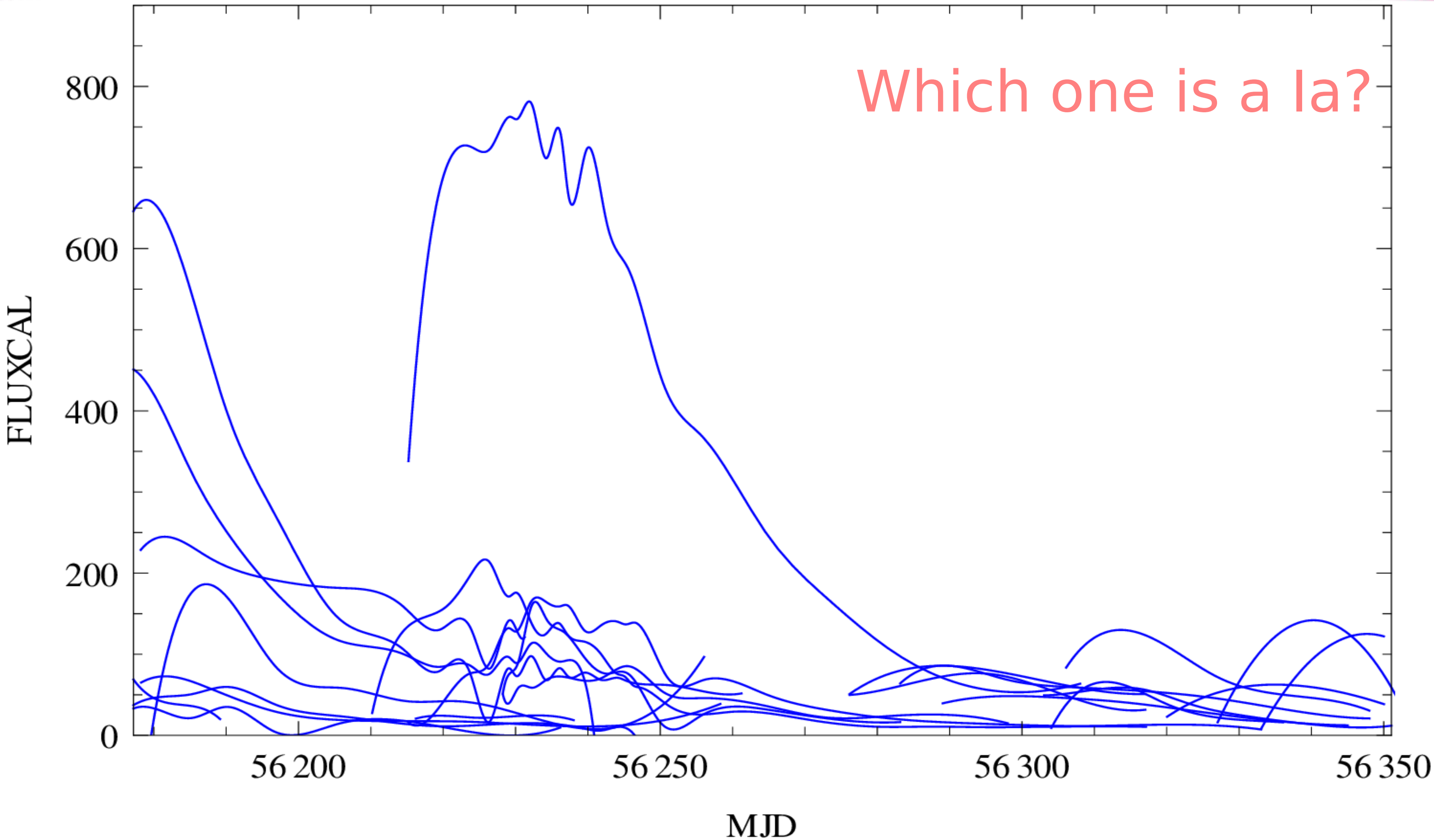
Type Ia SN surveys are not able to provide complete spectroscopic follow up

FAPESP

1. The problem

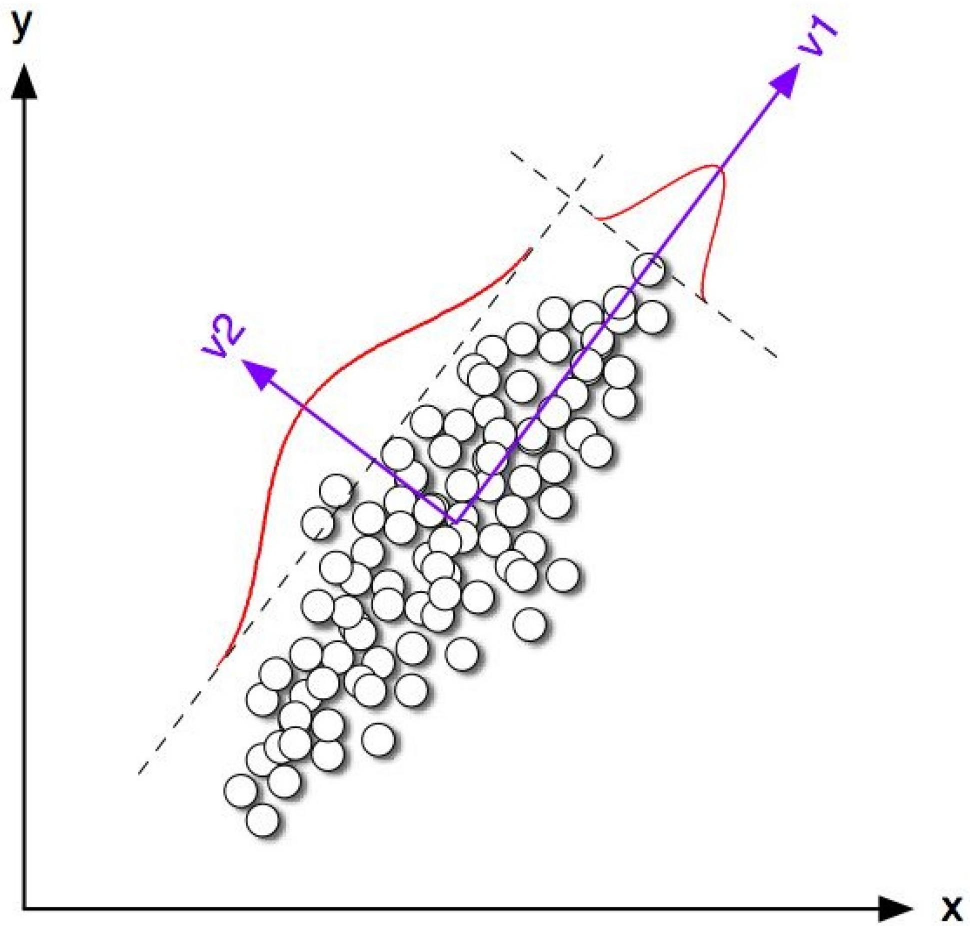


1. The problem



FAPESP

2. Principal Component Analysis (PCA)

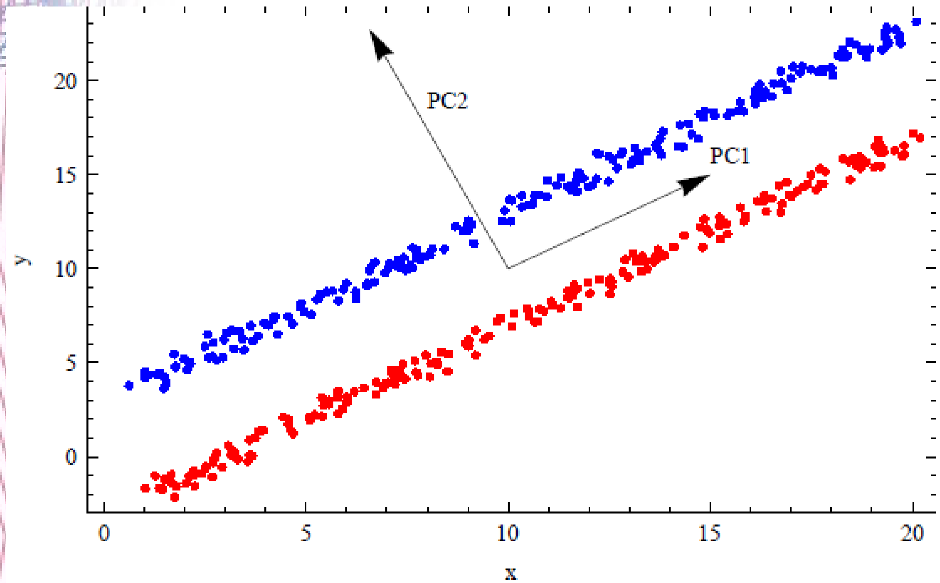


http://web.media.mit.edu/~tristan/phd/dissertation/figures/PCA.jpg

Dimensionality reduction technique

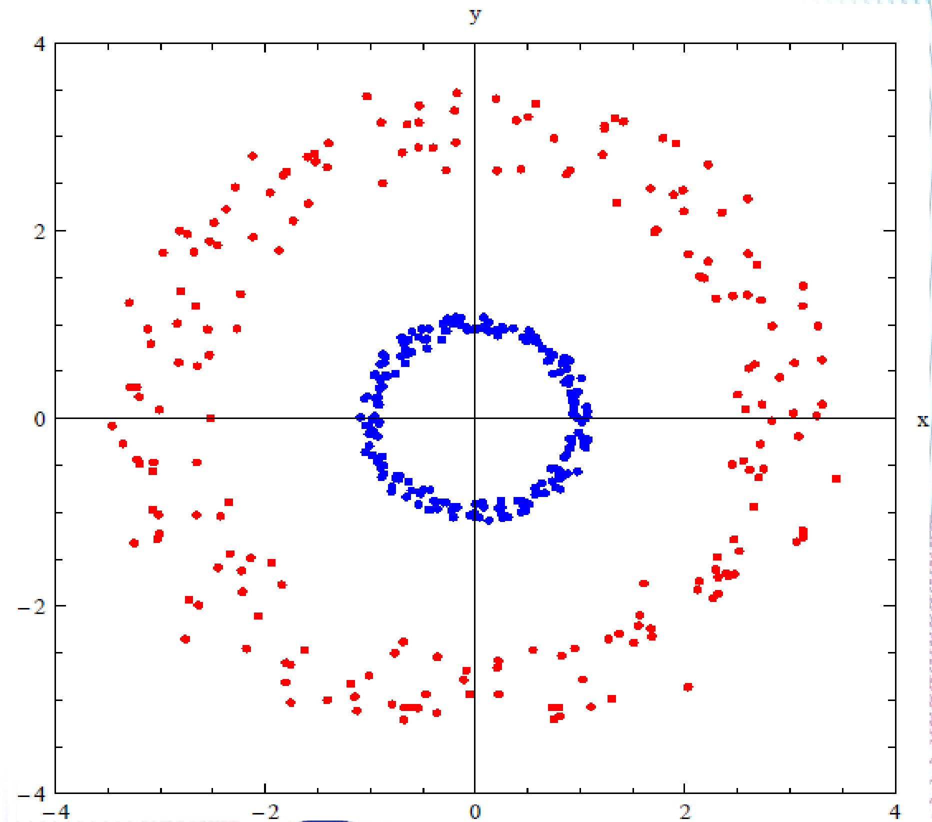
Look for directions that maximizes variance

2. PCA limitations

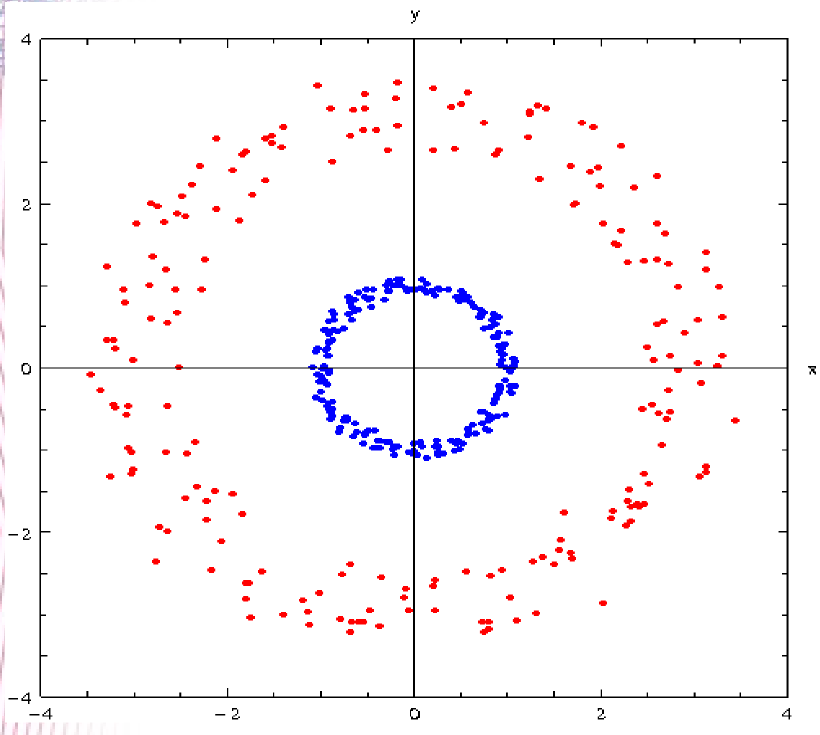


Is not designed to capture non-linear structure

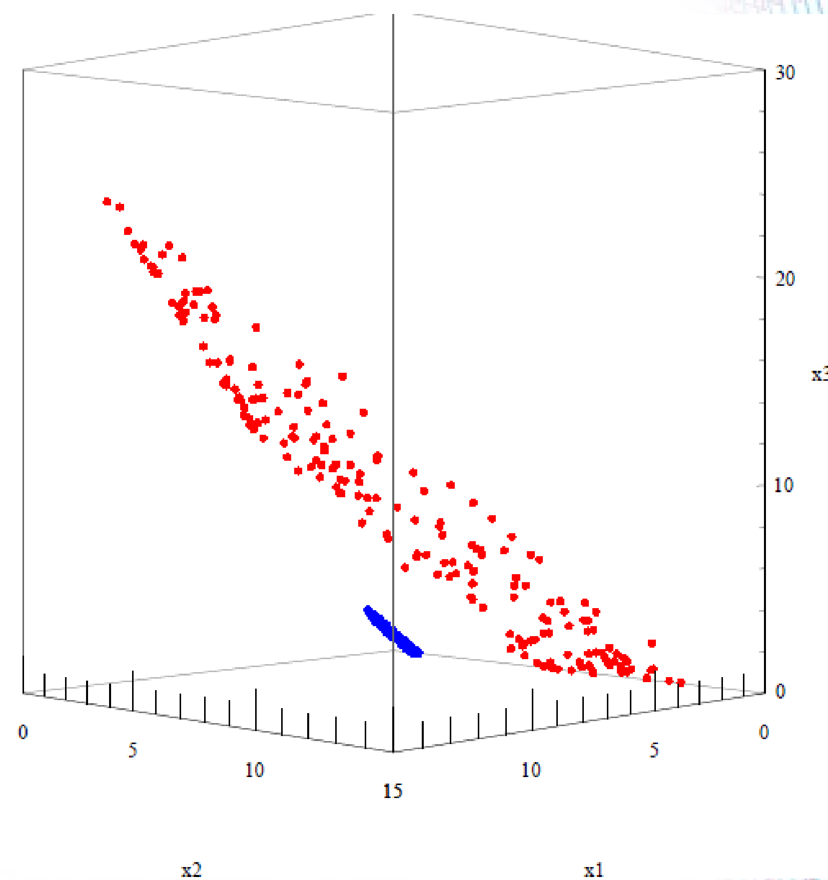
Does not care about labels



2. PCA extensions



$$(x, y) \quad \rightarrow \quad z = \sqrt{x^2 + y^2}$$



Sometimes, going to higher dimensions might solve the problem

2.The kernel trick

In the linear case, with

$\boldsymbol{x}_i \quad \rightarrow i-th$ data vector,

$$K_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$\alpha_k \rightarrow k-th$ eigenvalue

$\boldsymbol{v_k} \rightarrow k-th$ eigenvector

$$\boldsymbol{v}_k^T \boldsymbol{n} = \sum_{i=1}^{N} \alpha_k K(\boldsymbol{x_i}, \boldsymbol{n})$$

where $\boldsymbol{n}$ is the data to be projected and

$$K(\boldsymbol{x}_i, \boldsymbol{n}) = \boldsymbol{x}_i^T \boldsymbol{n}$$

In the kernel approach,

$\Phi : \mathbb{R} \rightarrow \mathbb{F}$

$\boldsymbol{x} \rightarrow \Phi(\boldsymbol{x})$

$$K_F(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j)$$

$$\boldsymbol{v}_\Phi^l \cdot \Phi(\boldsymbol{n}) = \sum_{i=1}^{N} \alpha_{\Phi_i}^l K_F(\boldsymbol{x}_i, \boldsymbol{n})$$

First natural choice - Gaussian kernel

$$k(\mathrm{x}_i, \mathrm{x}_j) = \exp\left[\frac{\|\mathrm{x}_i - \mathrm{x}_j\|^2}{2\sigma^2}\right]$$

2.The kernel trick

In the linear case, with

$\boldsymbol{x}_i \quad \rightarrow i-th$ data vector,

In the kernel approach,

$\Phi : \mathbb{R} \rightarrow \mathbb{F}$

It is not necesary to know the mapping!

where n is the data

to be projected and

$$K(\boldsymbol{x}_i, \boldsymbol{n}) = \boldsymbol{x}_i^T \boldsymbol{n}$$

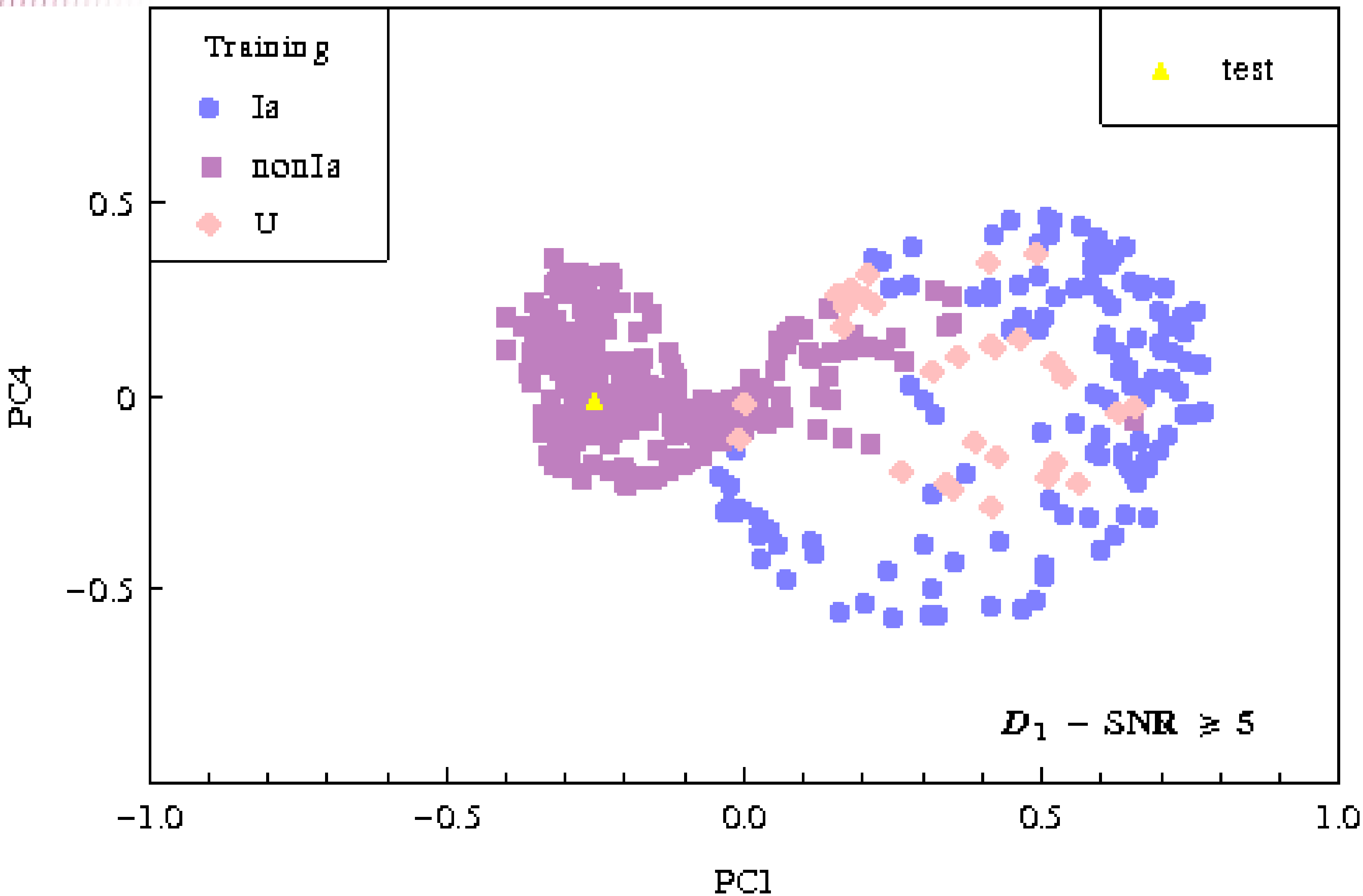
First natural choice - Gaussian kernel

$$k(\mathrm{x}_i, \mathrm{x}_j) = \exp\left[\frac{\|\mathrm{x}_i - \mathrm{x}_j\|^2}{2\sigma^2}\right]$$

FAPESP

3. kPCA applied to SNe classification

Classification

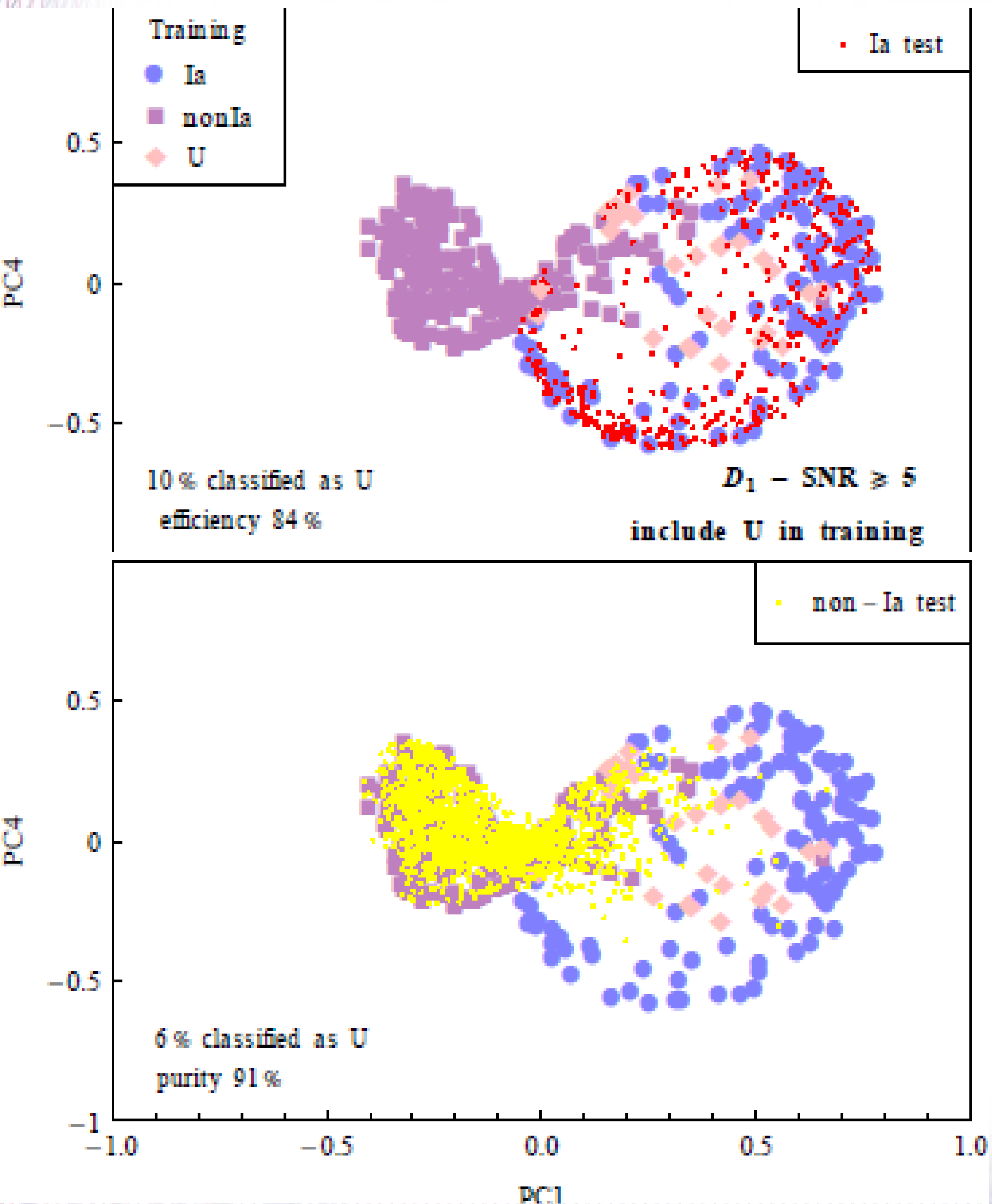


THE
Nearest
Neighbor

(1NN)

FAPESP

3. kPCA applied to SNe classification



Post-SNPCC sample
After selection cuts

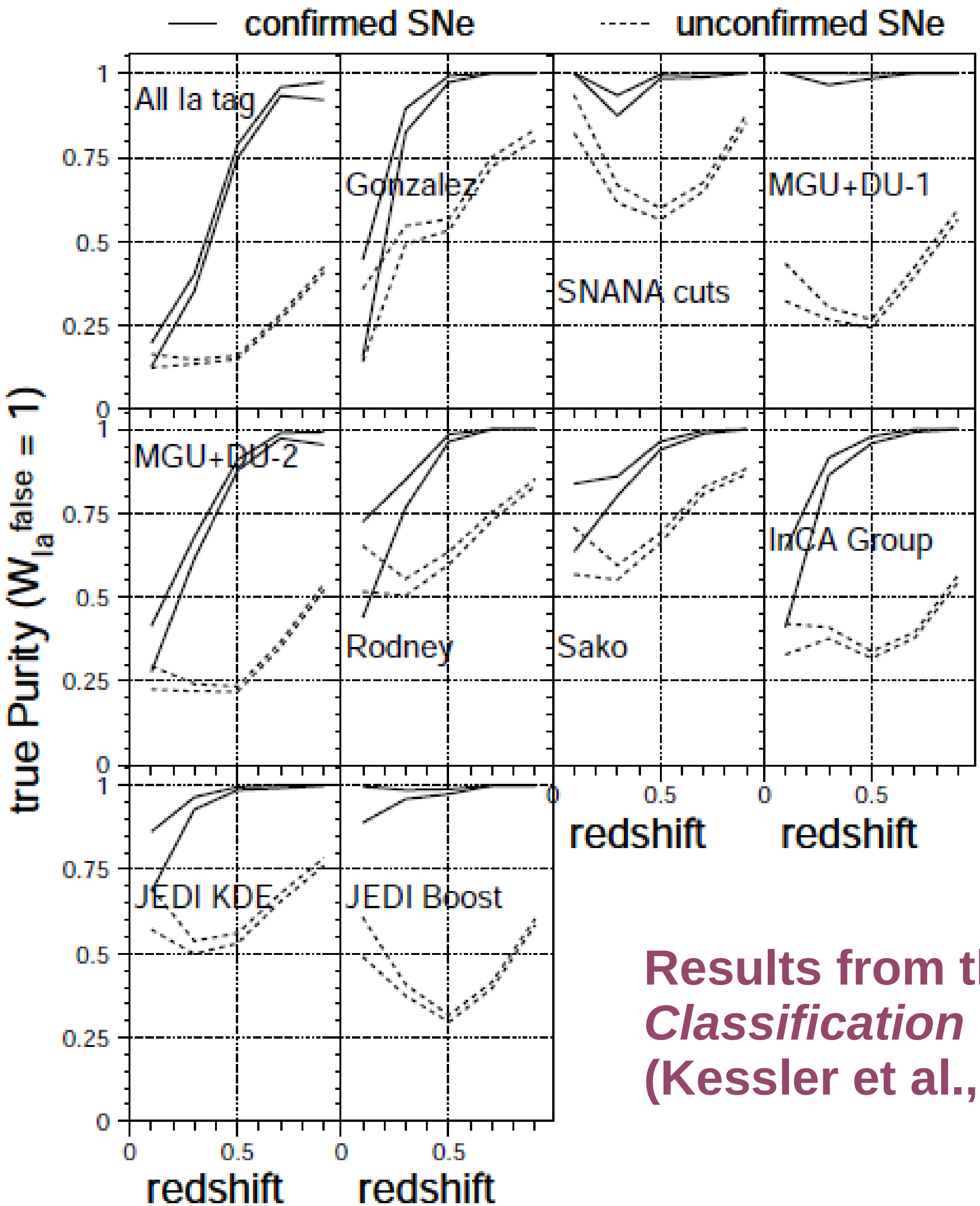
FoM ~0.60

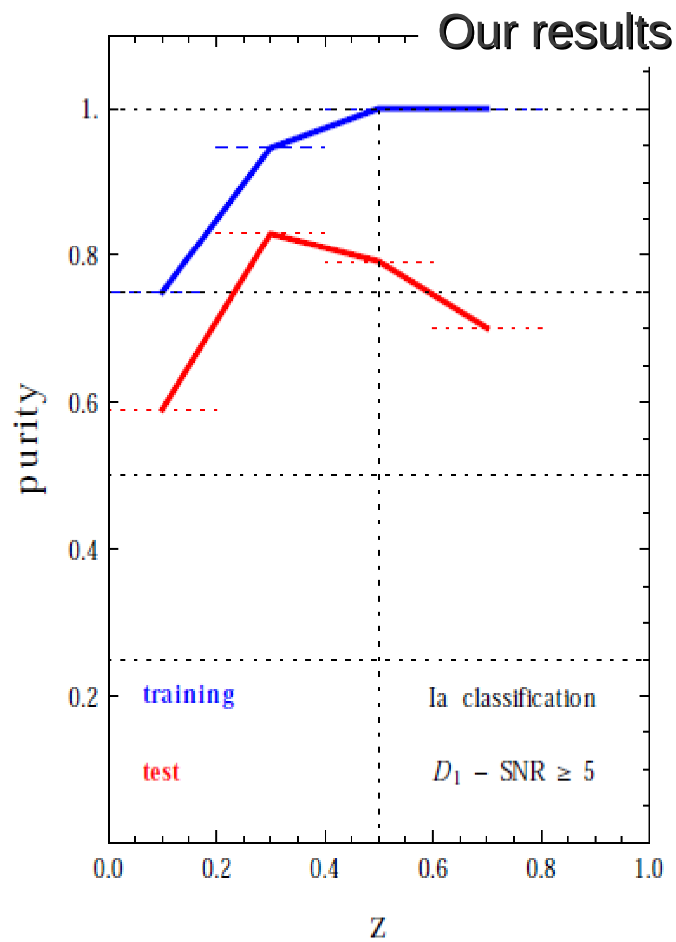
SC ~91%

<u>Selection cuts:</u>
{-3,+24} in r-band
At least 3 obs with SNR>5
in each band

FAPESP

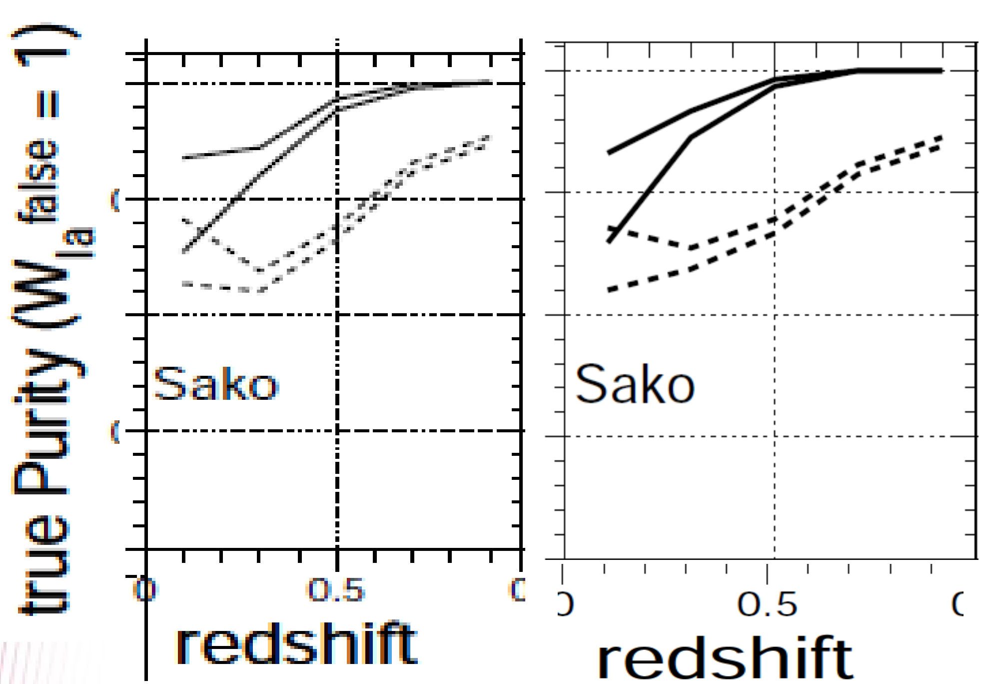
3. kPCA applied to SNe classification





Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

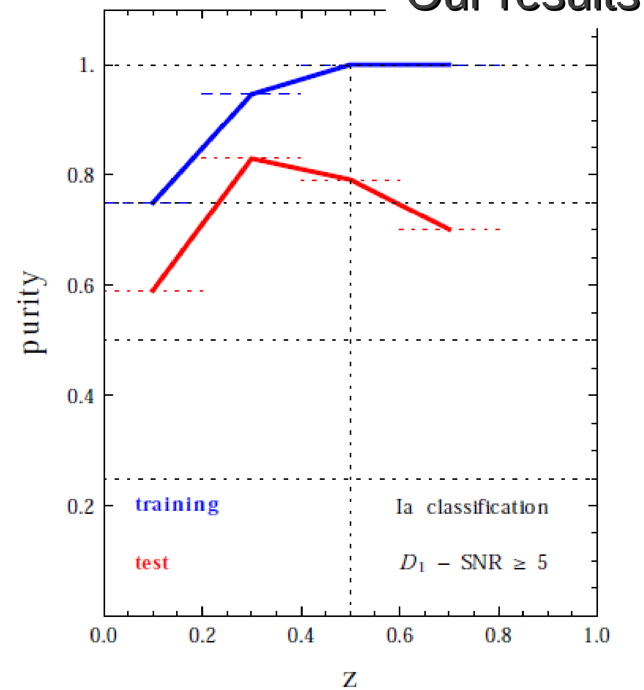
3. kPCA applied to SNe classification



No hostZ

hostZ

Our results



average purity: 75%

Same result as winner SNPCC (76%), without using host redshift information

Results from the *Supernova Photometric Classification Challenge* (Kessler et al., 2010)

Better result in intermediate redshift

FAPESP

4. Conclusions

1. SNe photometric classification is not a future issue..
   it is already here!

2. kPCA is a powerfull tool, mainly if we are interested in a high quality purity in intermediate redshifts.

3. There is no need of enviromental, redshift or astrophysical hypothesis

4. Great potential in detecting previously non-observed objects: Application to PISN search

FAPESP