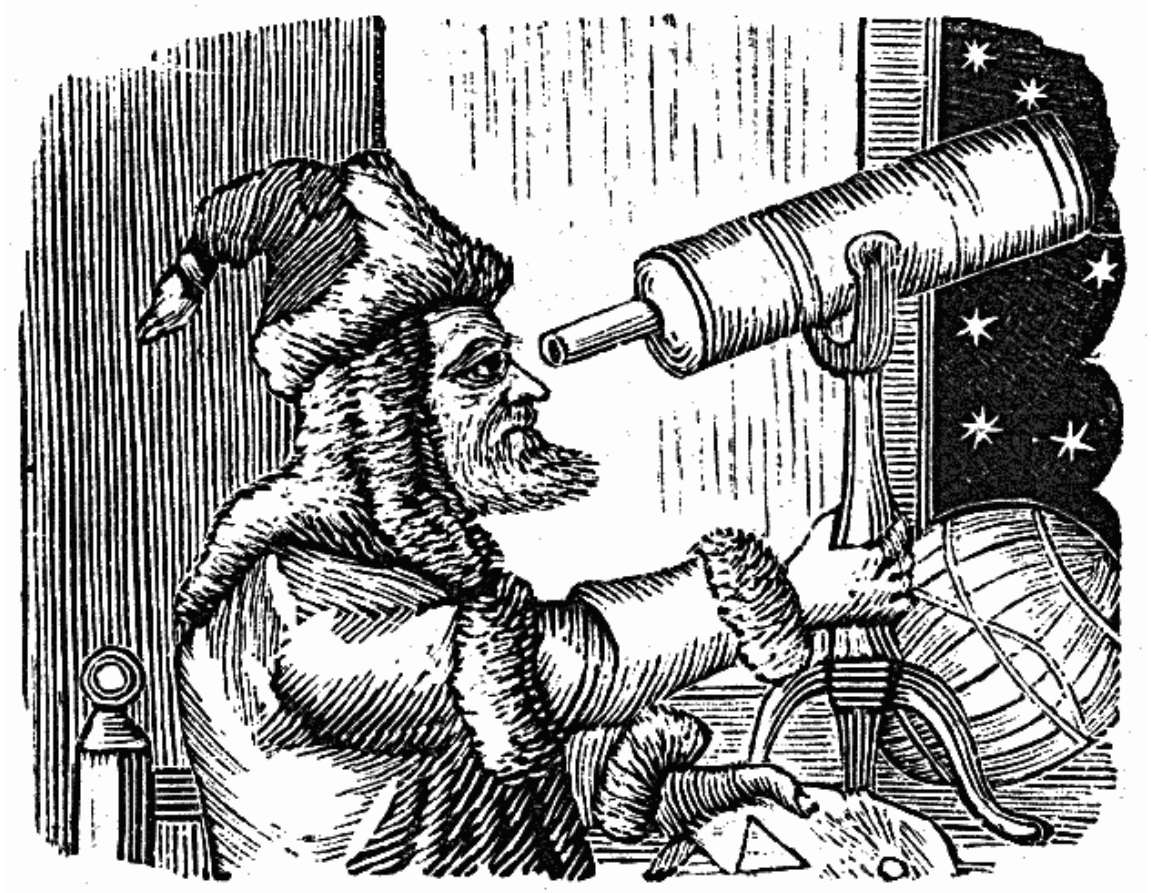
Mathematical characterization of galaxies using IFS data: I - analysis of spectral information

Yago Ascasibar
(Universidad Autónoma de Madrid)

INAOE, 01/09/2014

Outline



The traditional approach

Characterization of spectra

Physical modelling

- Mean stellar age/SFH
- Stellar/gas metallicity
- Ionization parameter
- Presence/contribution of AGN/shocks/...
- ...

Feature extraction

- Broadband magnitudes/colours
- Emission lines/ratios
- Lick indices
- Redshift/velocity/moments
- ...

The astrostatistical approach

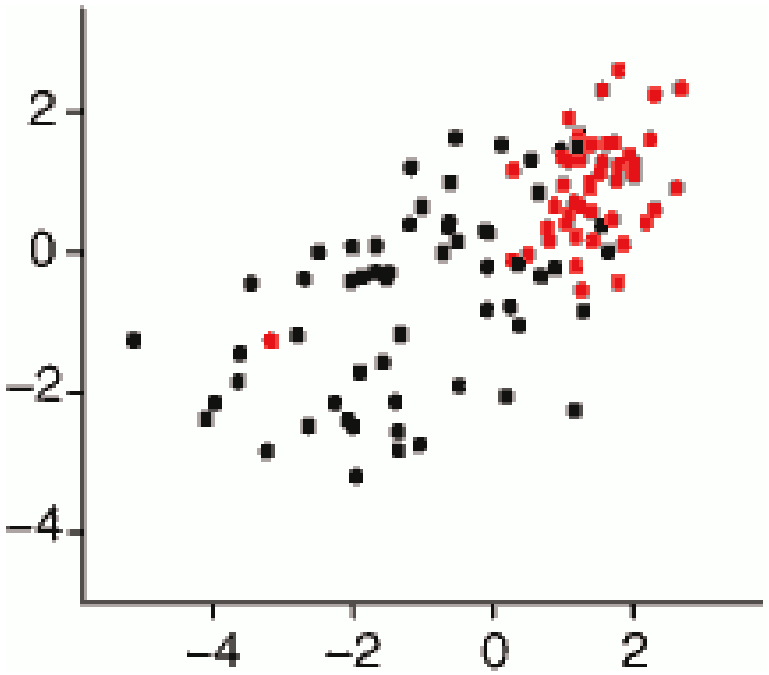
Spectra as multi-dimensional data

How many spatial dimensions?

- Fiber: 0
- Long-slit: 1
- IFS: 2

How many spectral dimensions?

- One single spectrum: N_λ



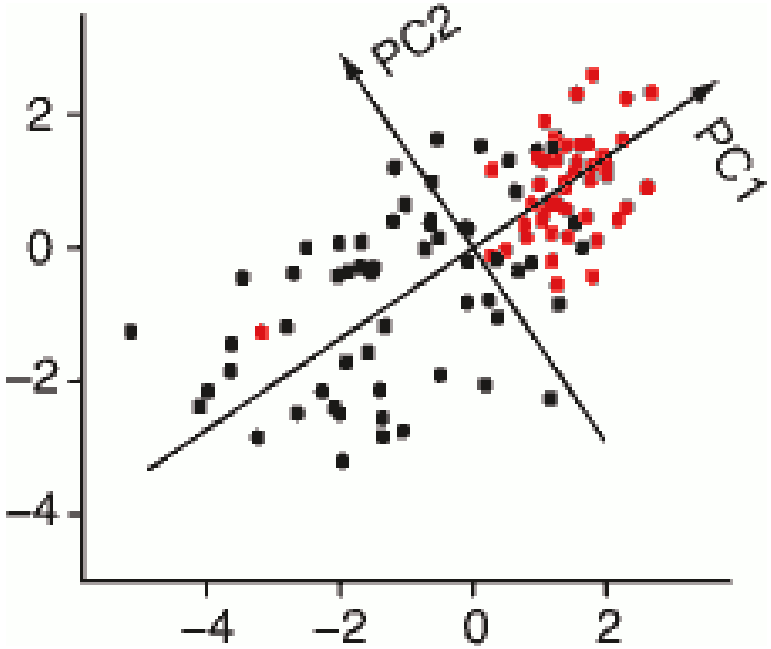
Principal Component Analysis (PCA)

Find "principal axes" of the distribution

1. Compute centre
2. Compute covariance matrix
3. Find (largest) orthogonal eigenvalues/vectors

What for?

- Dimensionality reduction
- Physical interpretation?



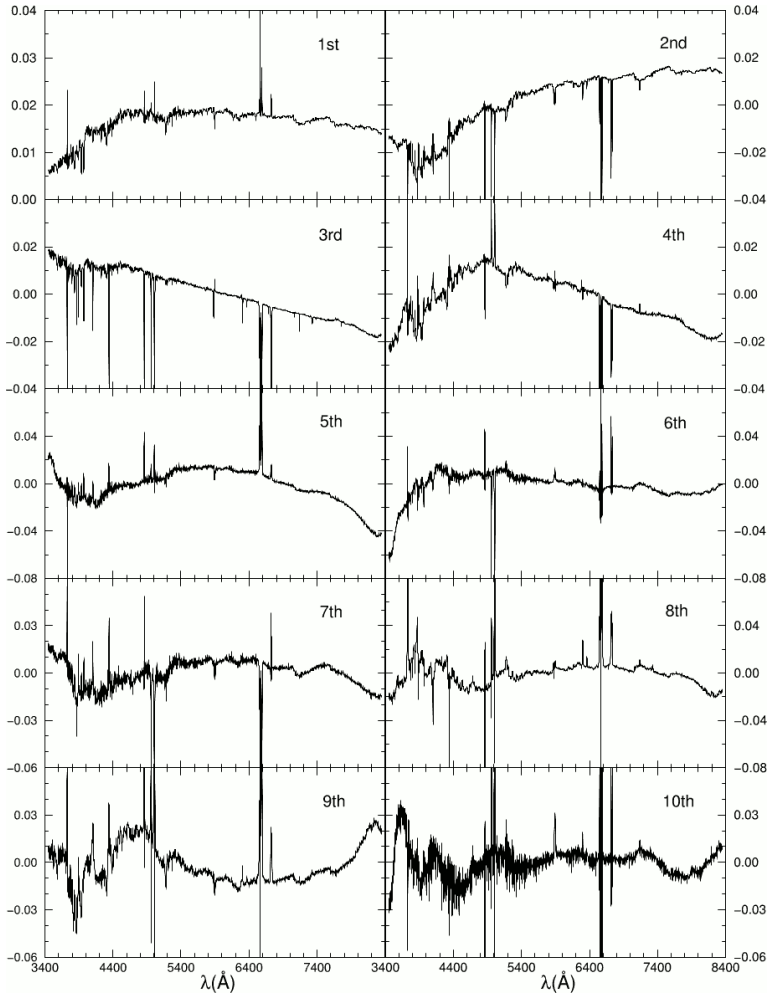
Principal Component Analysis (PCA)

Find "principal axes" of the distribution

1. Compute centre
2. Compute covariance matrix
3. Find (largest) orthogonal eigenvalues/vectors

What for?

- Dimensionality reduction
- Physical interpretation?



Yip et al. (2004)

Principal Component Analysis (PCA)

Advantages

- Simple
- Fast
- Scalable

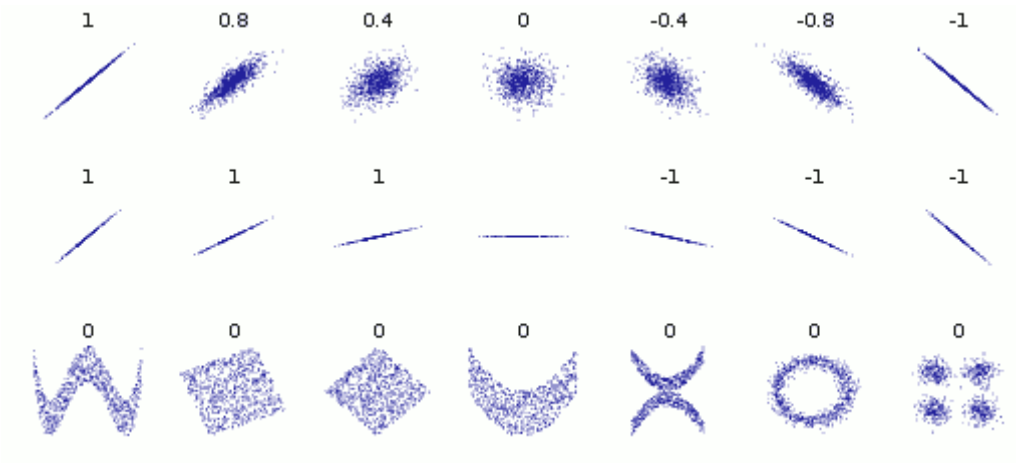
Disadvantages

- Uncorrelated $\neq$ independent
- Non-linear manifolds
- Metric-dependent

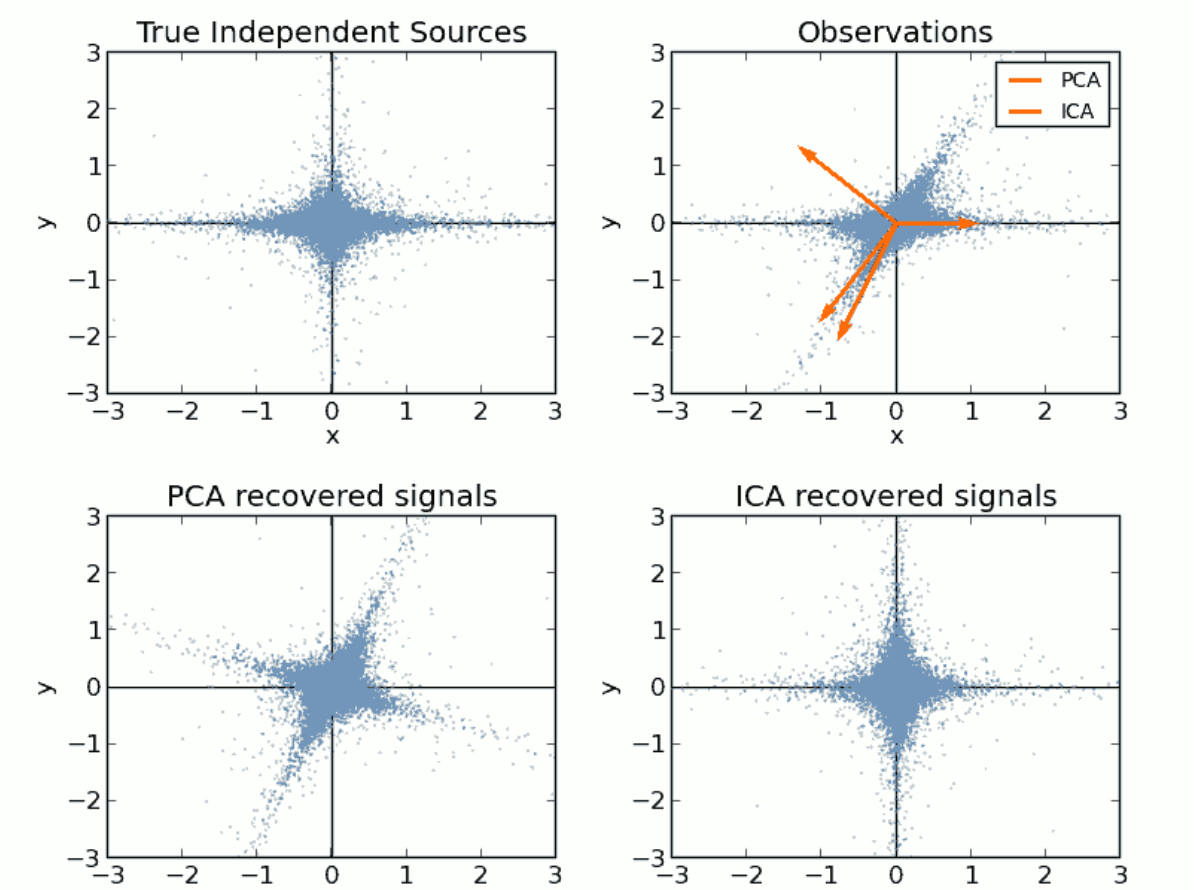
Independent Component Analysis (ICA)

x and y are...

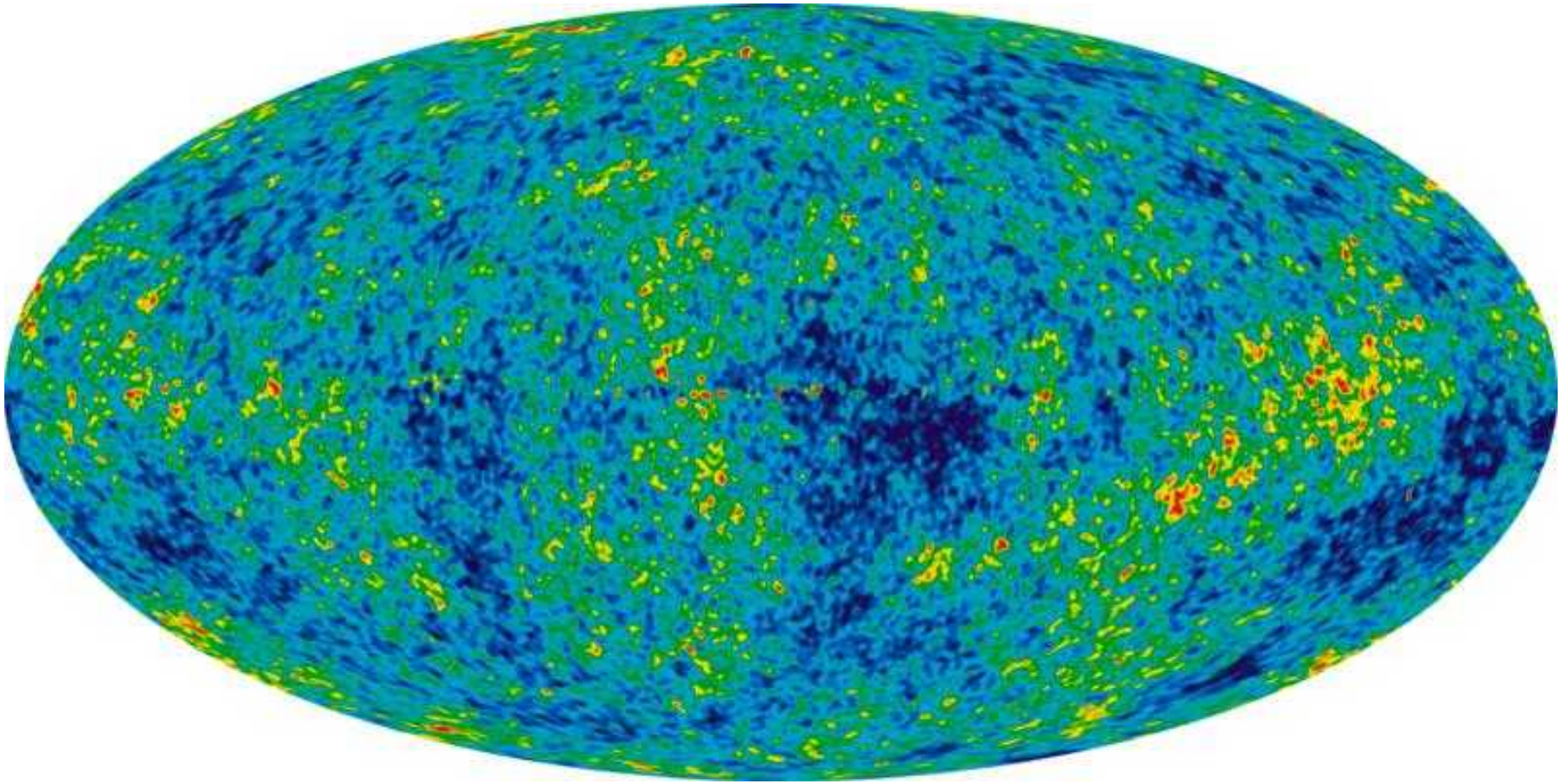
- uncorrelated $\Leftrightarrow \mathrm{Cov}(x, y) = 0$
- independent $\Leftrightarrow p(x, y) = p(x)p(y)$



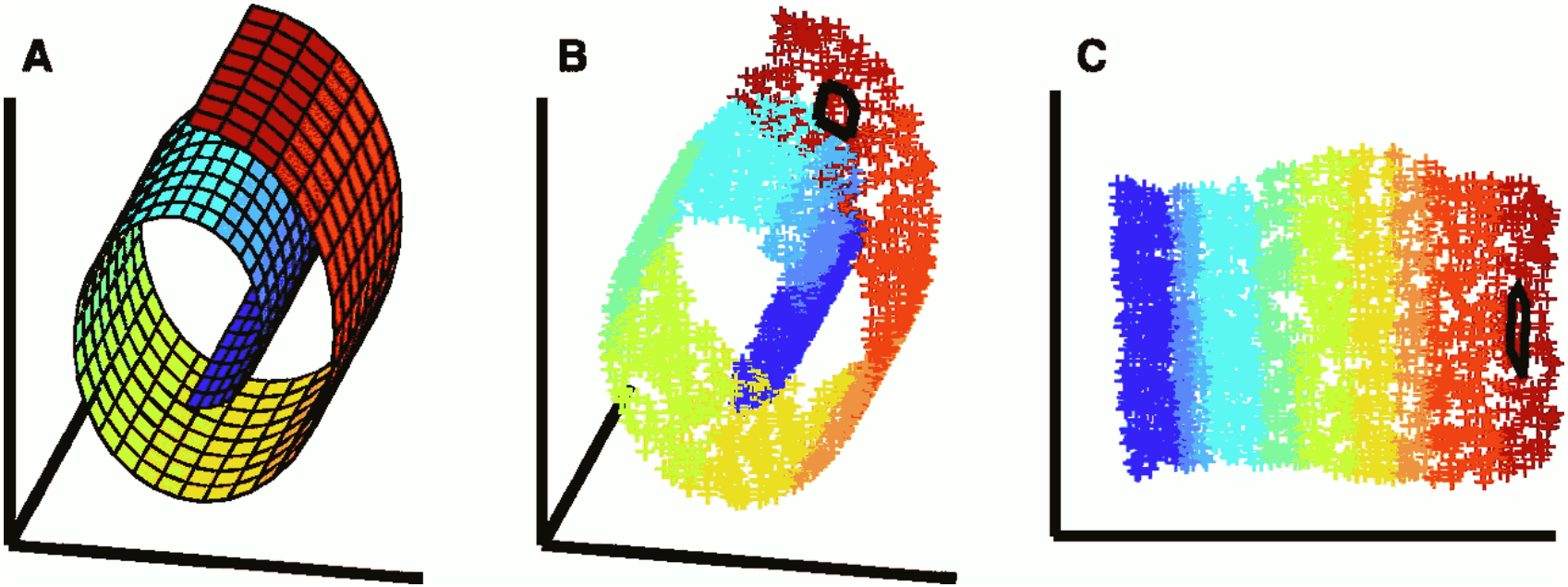
Independent Component Analysis (ICA)



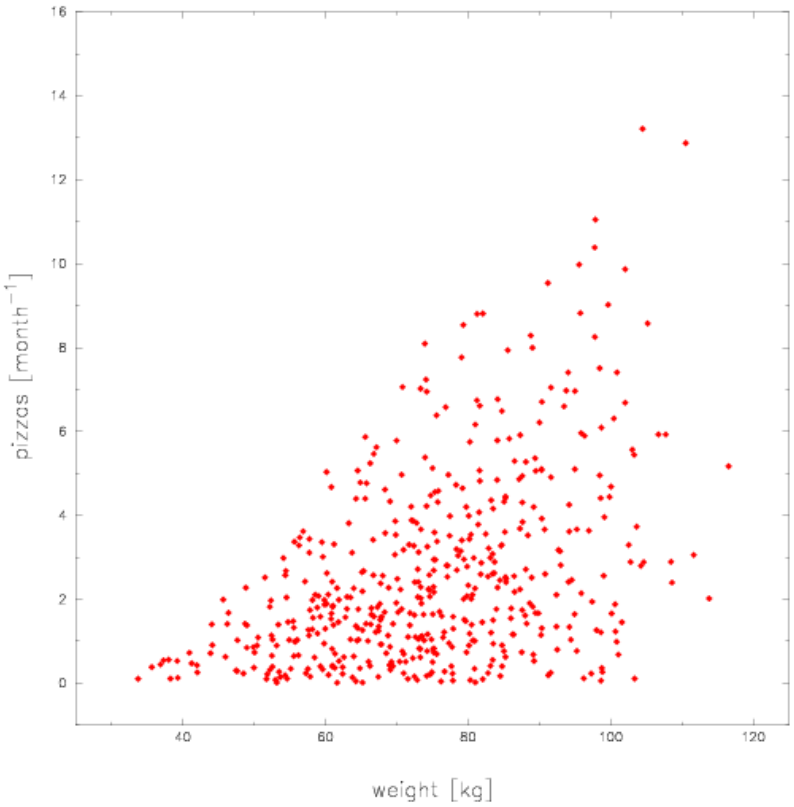
Independent Component Analysis (ICA)



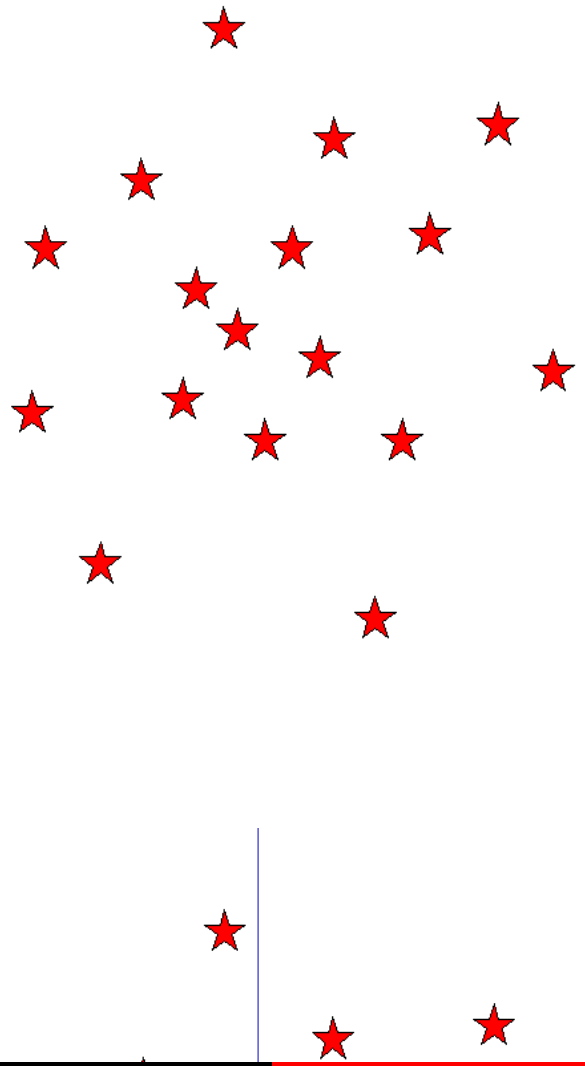
Manifold learning



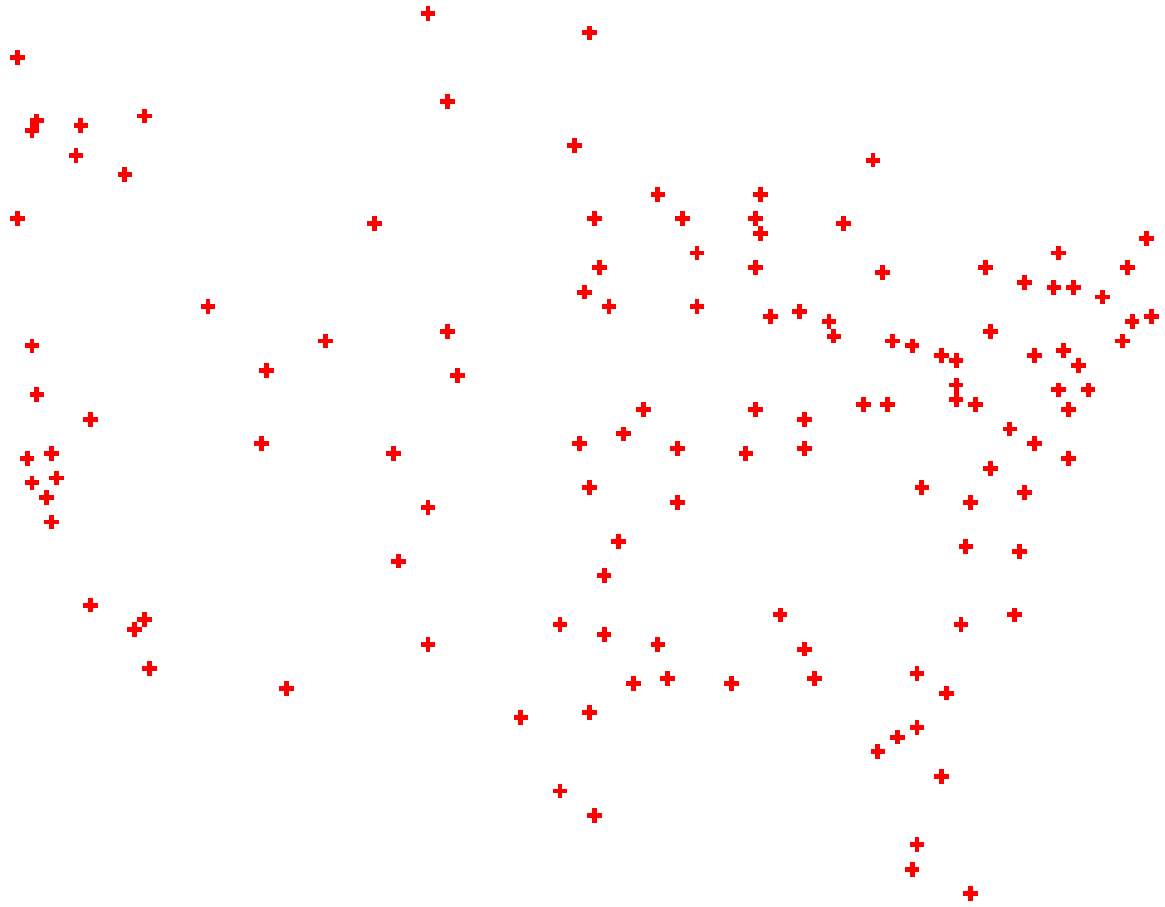
How to define a metric?



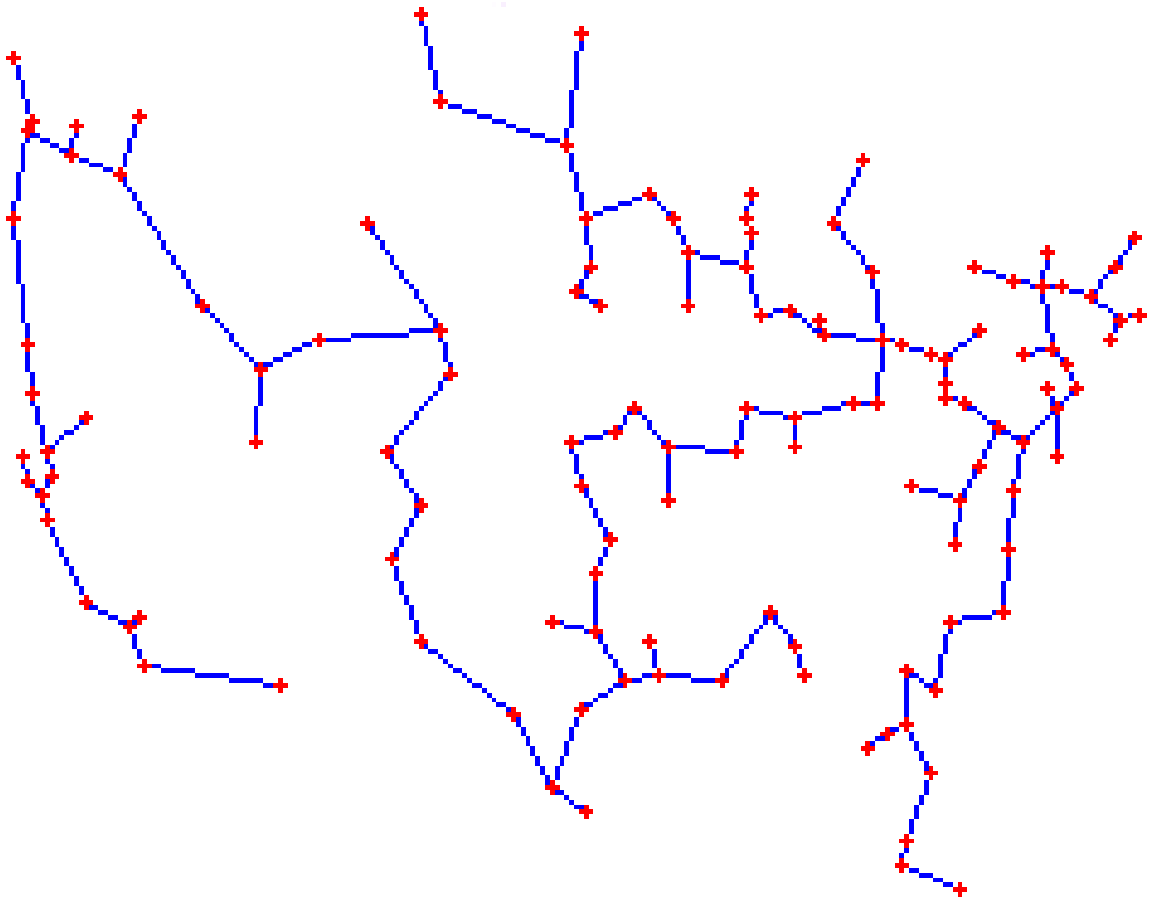
Field Estimator for Arbitrary Spaces



Minimal Spanning Tree (MST)



Minimal Spanning Tree (MST)

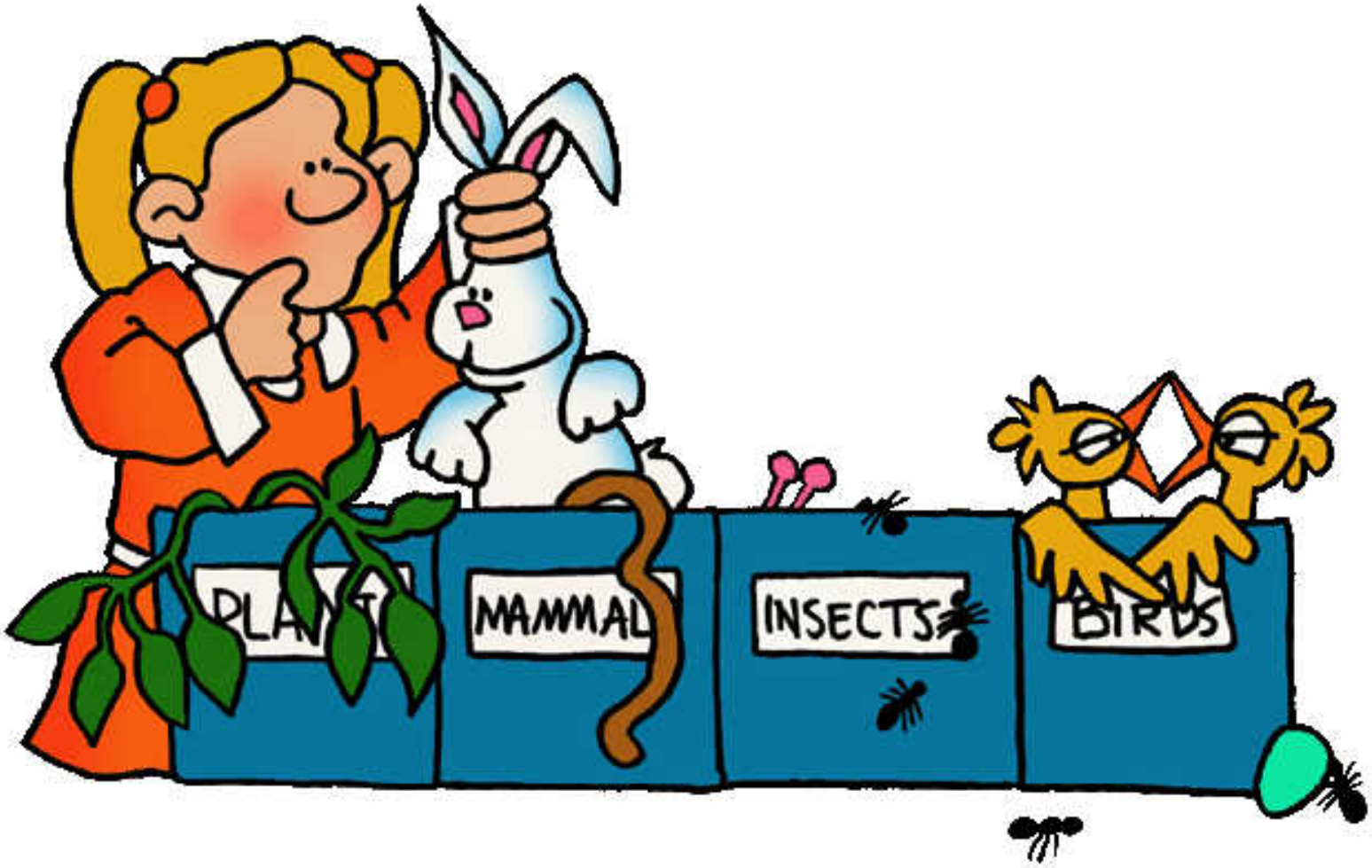


Summary

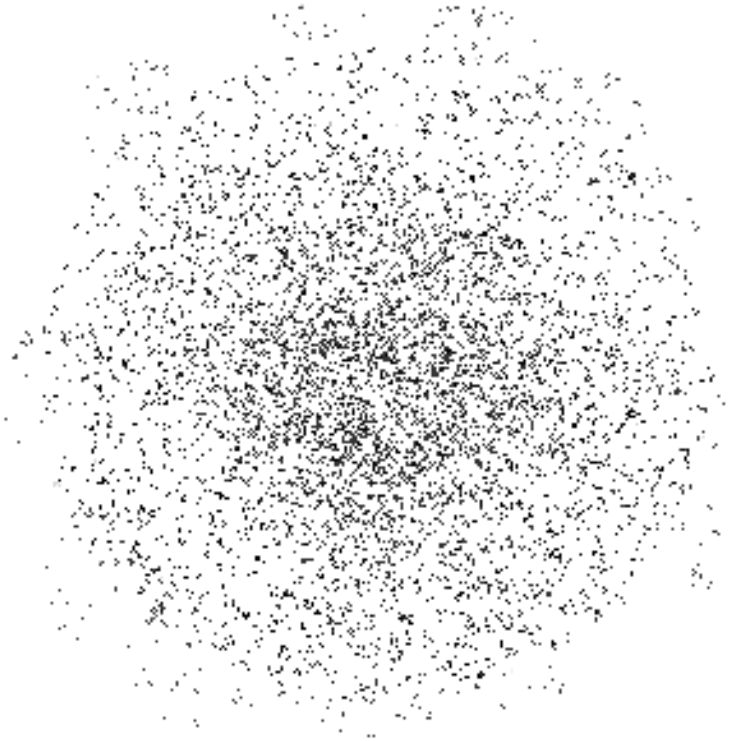
What to take home?

- A spectrum is a N_λ-dimensional vector
- Global distribution (e.g. PCA, ICA)
- Local distribution (e.g. FiEstAS, MST)

Classification



Classification



Classification

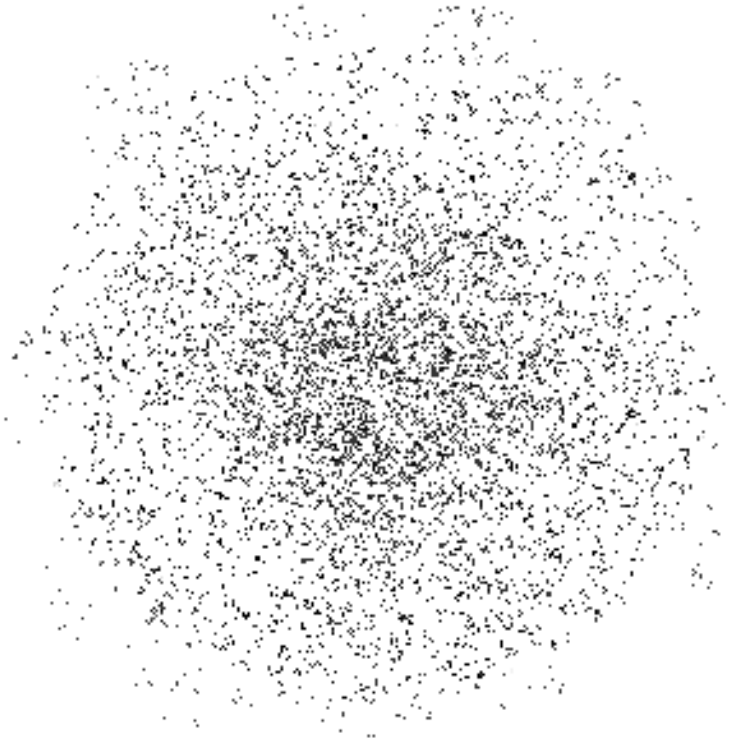
Supervised classification

- Nearest neighbour(s)
- Naive Bayes
- Decision trees/forests
- Support Vector Machines
- Artificial Neural Networks
- ...

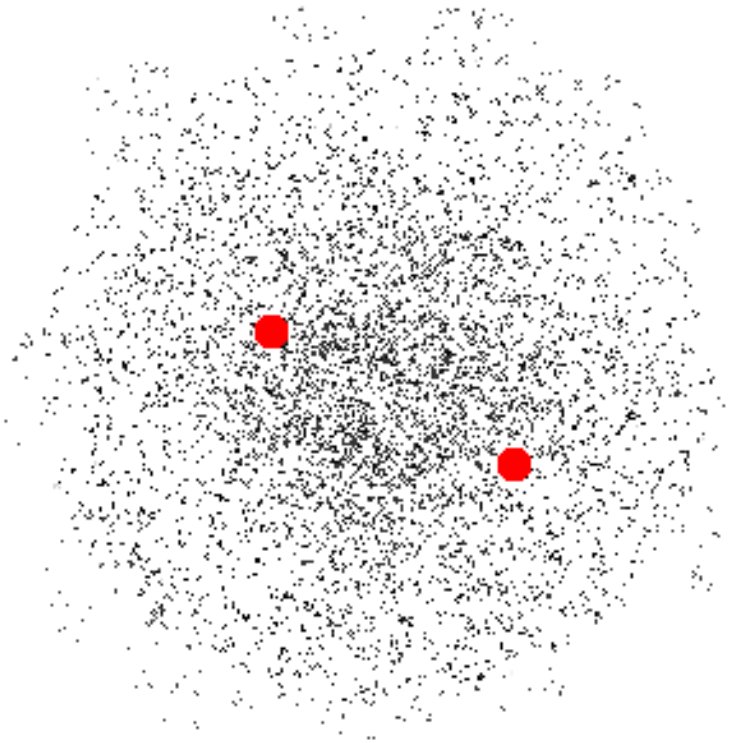
Unsupervised clustering

- ...
- k-means
- Expectation-Maximization
- HOT+FiEstAS

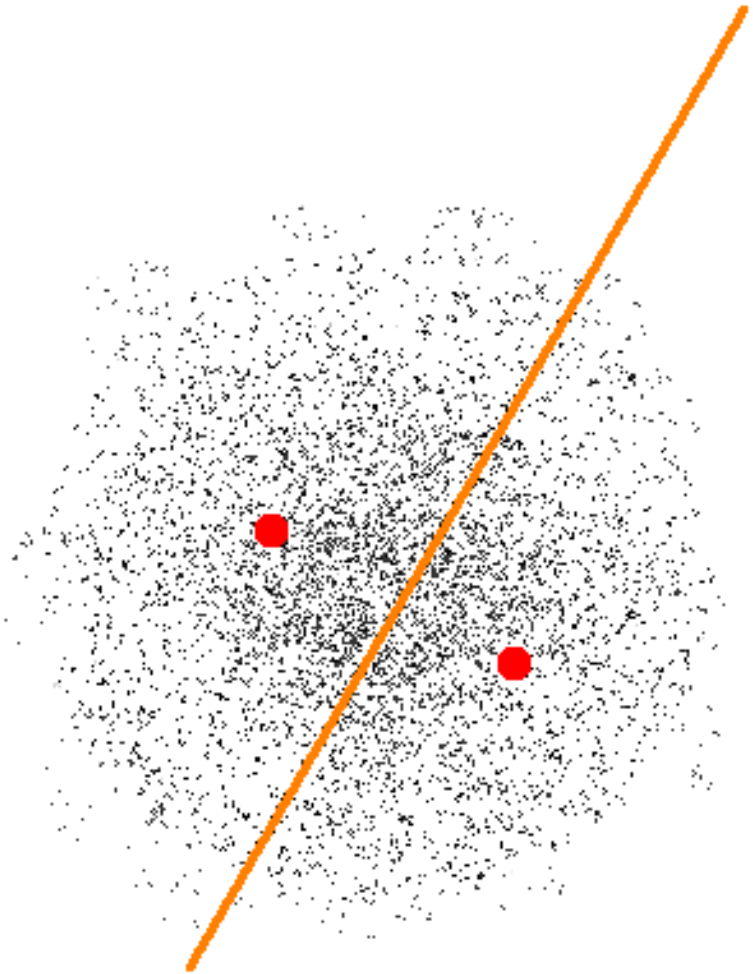
k-means



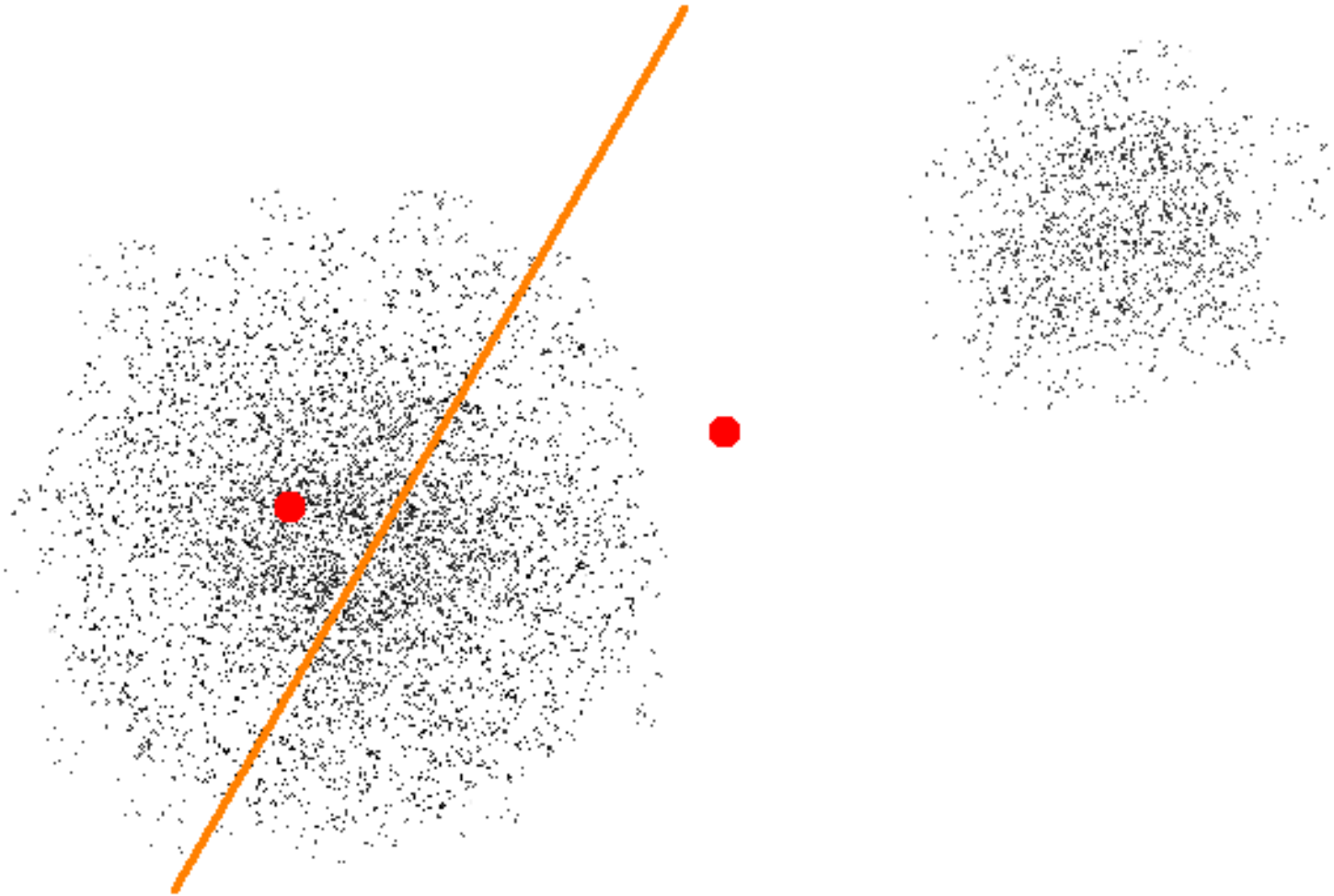
k-means



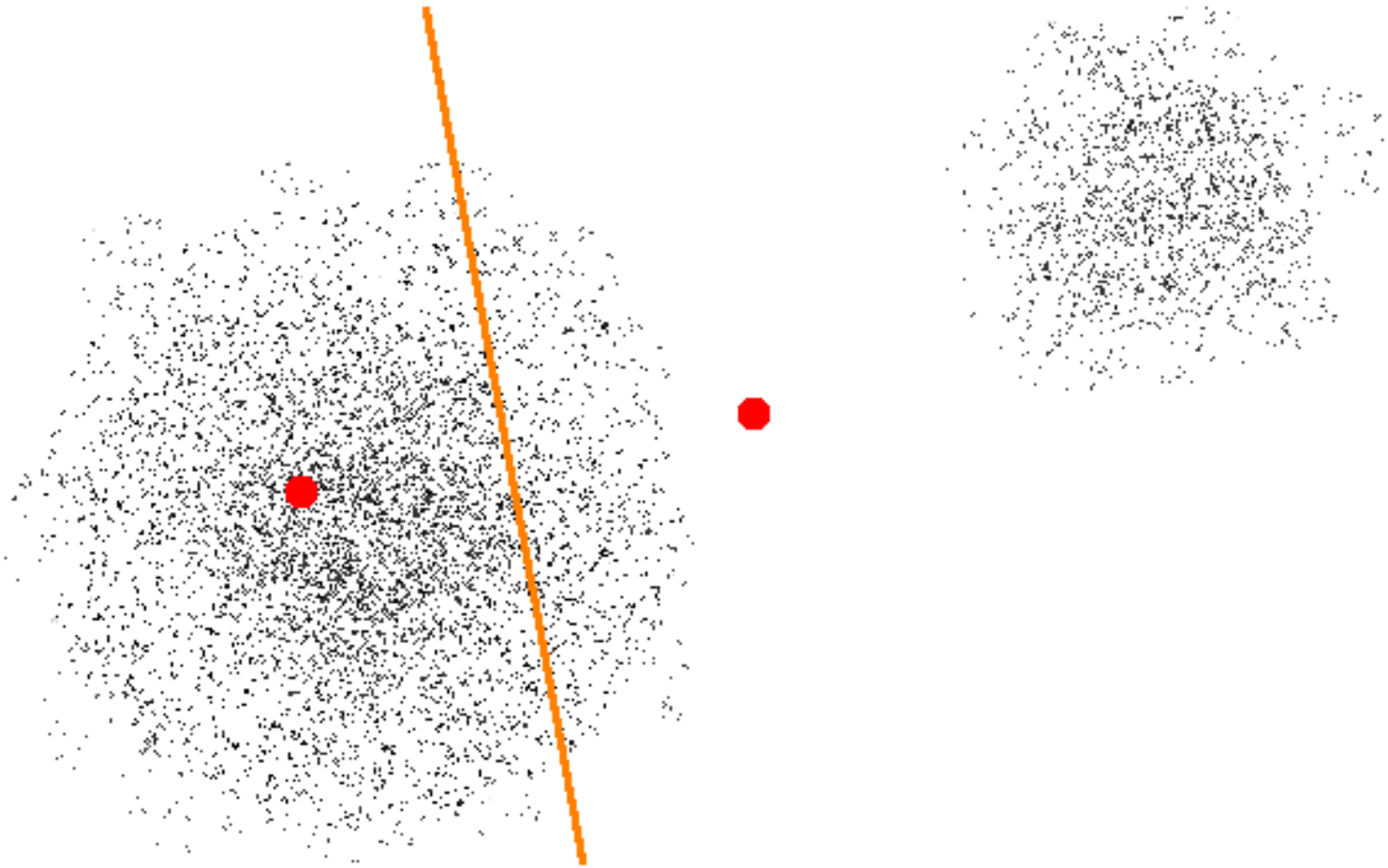
k-means



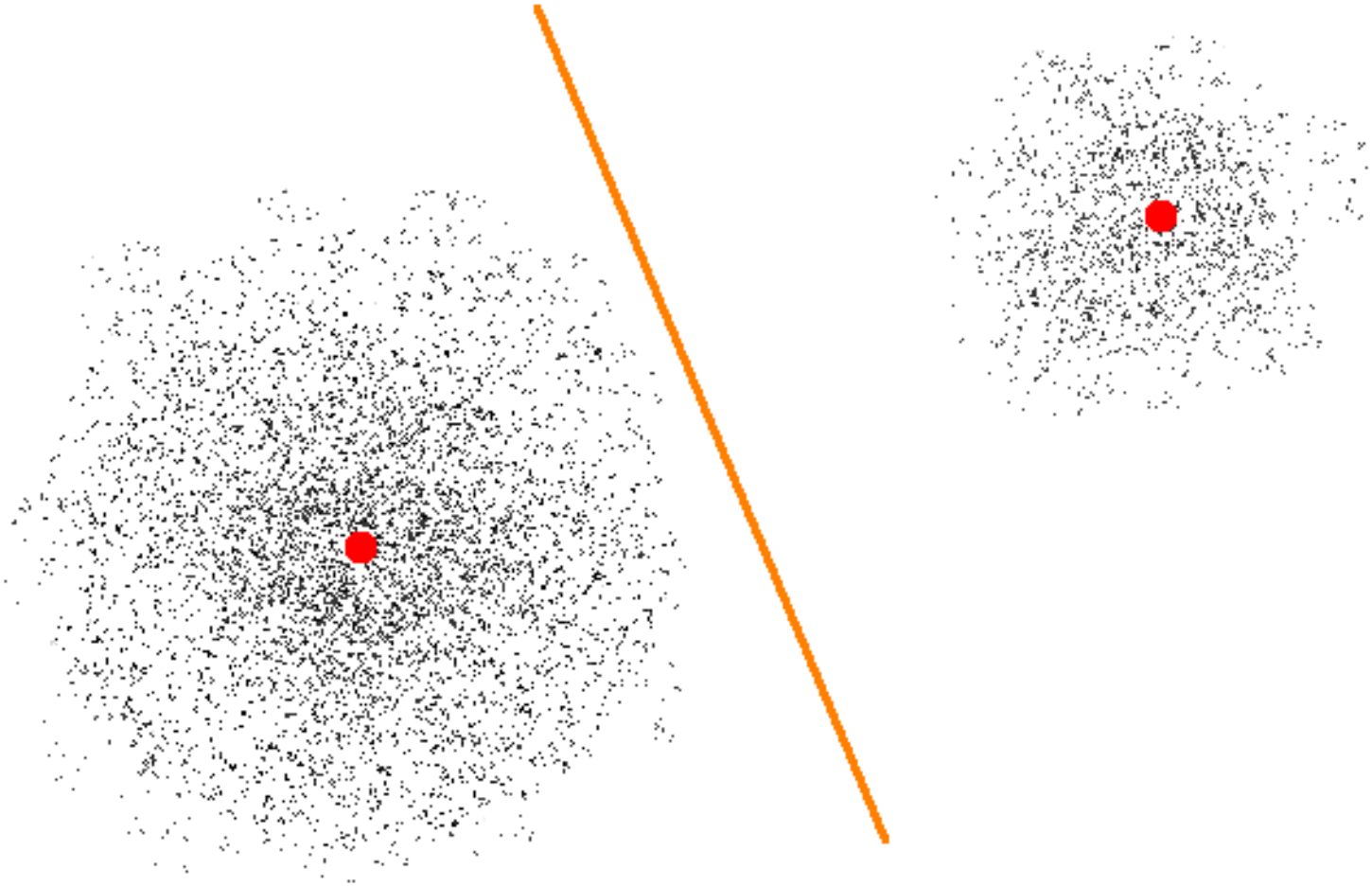
k-means



k-means



k-means



k-means

Advantages

- Simple
- Fast
- Scalable

Disadvantages

- Metric-dependent
- Initialization
  - number of classes
  - random seed
- Similar class sizes

Expectation-Maximization

Hard classification

$$\rho_A(x) > \rho_B(x) \Rightarrow x \in A$$

Soft classification

$$p(x \in A) = \frac{\rho_A(x)}{\rho_A(x) + \rho_B(x)}$$

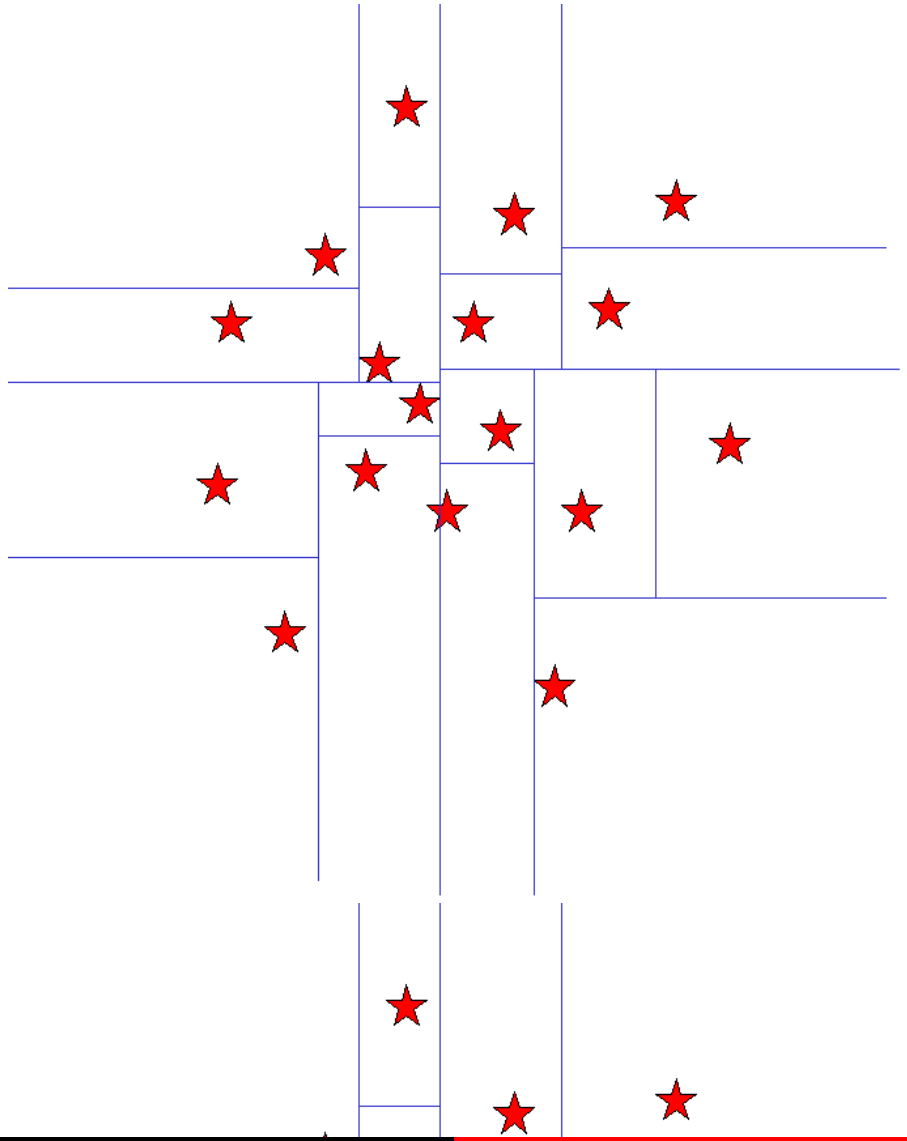
Gaussian mixture model

$$\rho_A(x) = \frac{N_A}{\sqrt{2\pi\sigma_A^2}} \, e^{-\frac{(x-x_A)^2}{2\sigma_A^2}}$$

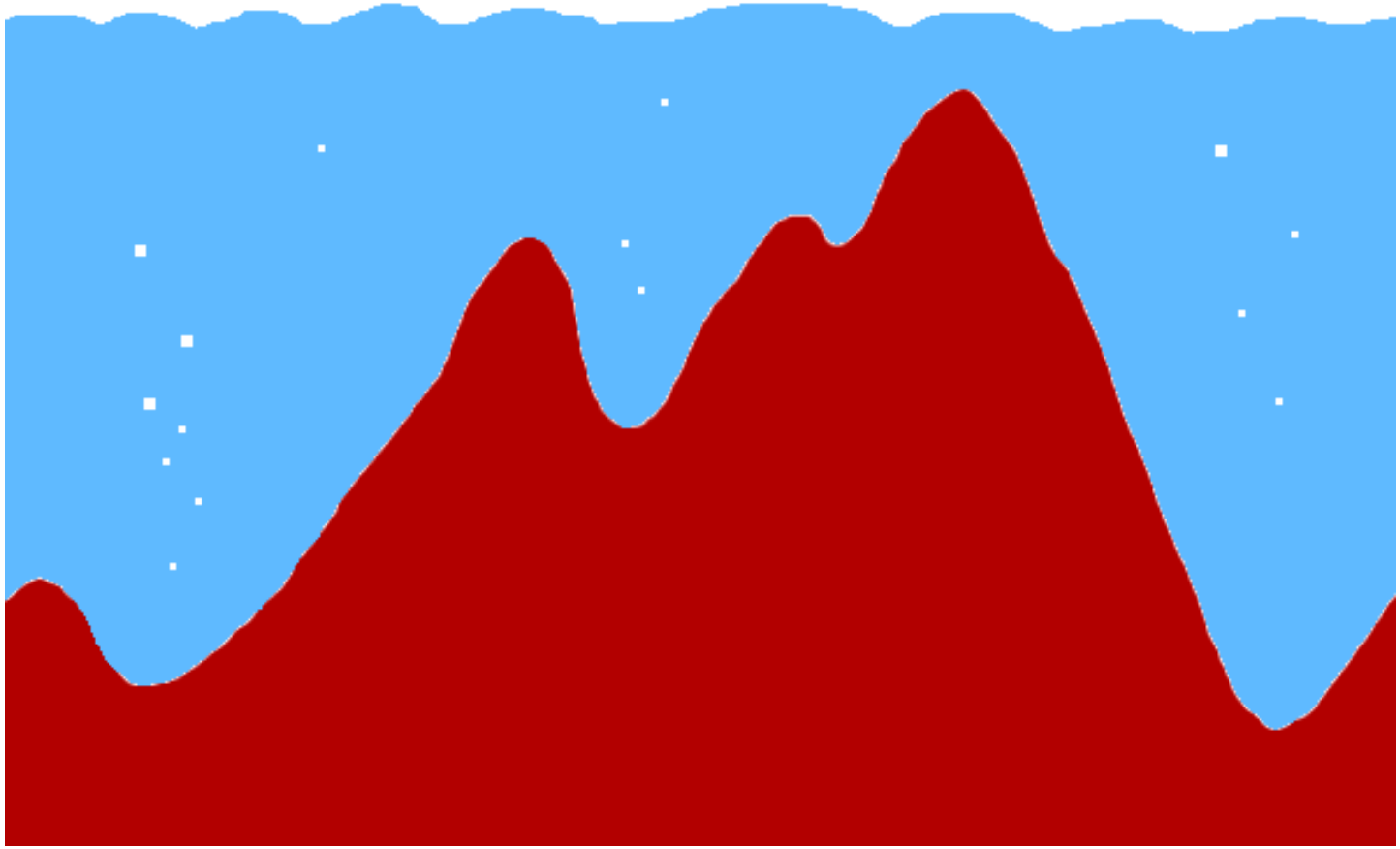
E-M algorithm

1. Estimate class parameters
2. Perform soft assignments
3. Iterate to convergence

Kernel density estimation



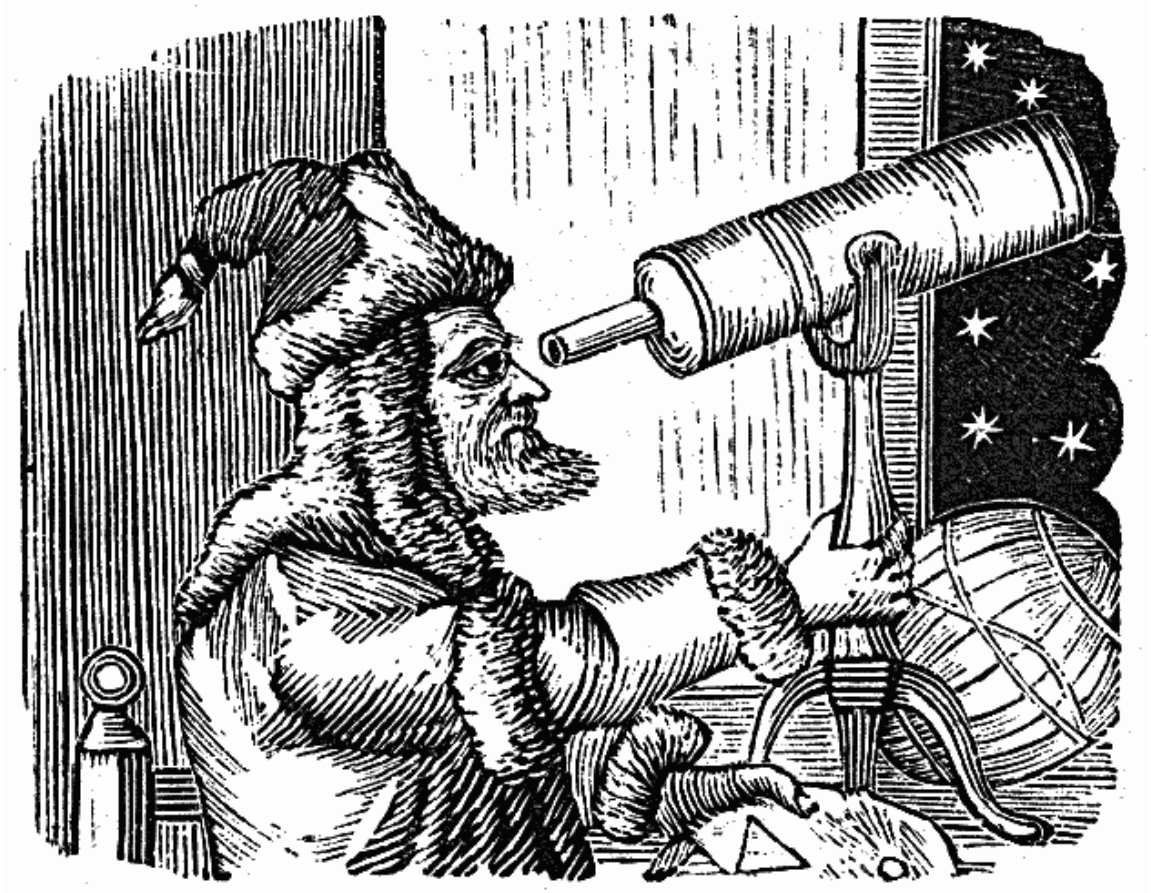
Hierarchical Overdensity Tree



Summary

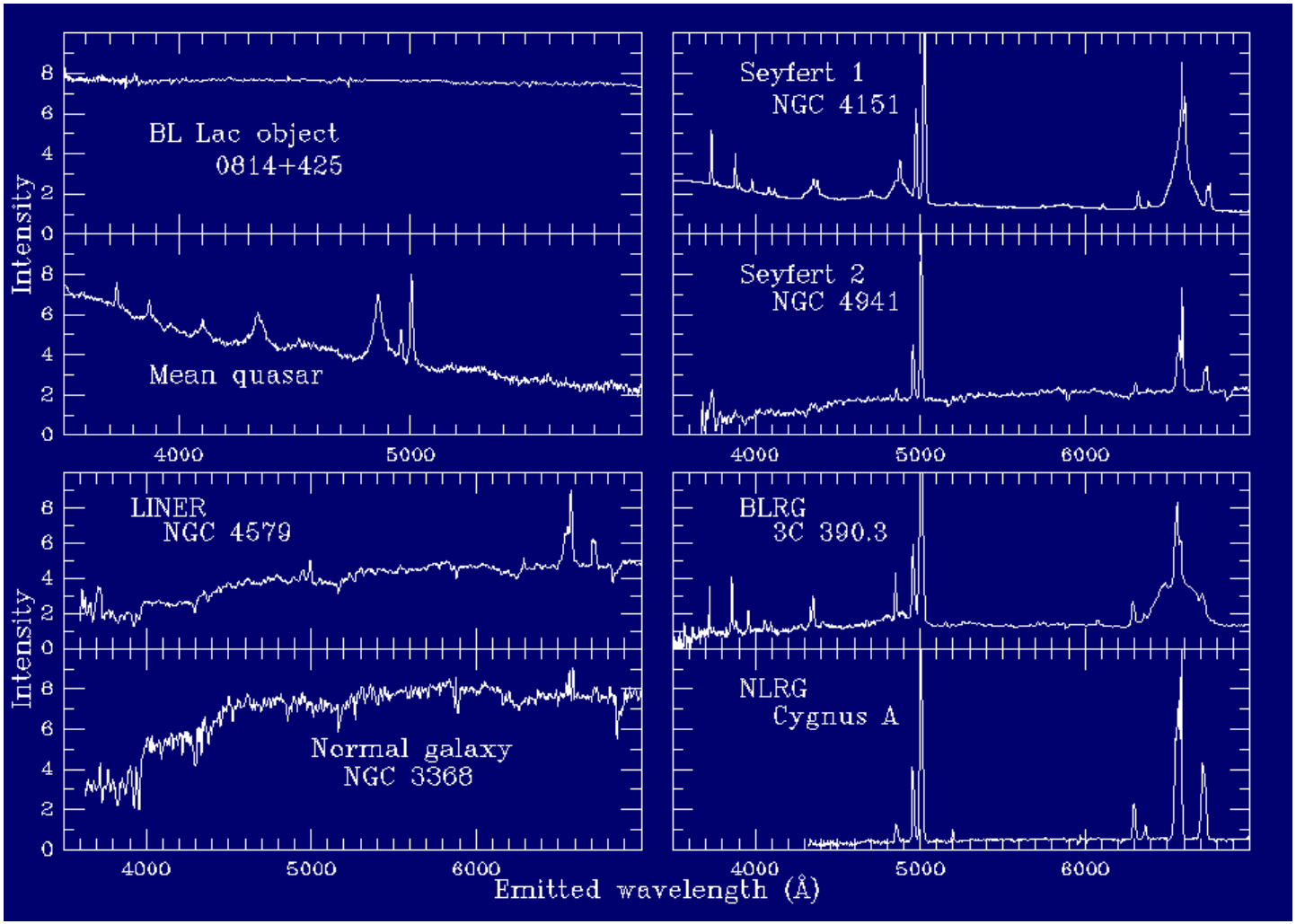
Classification

- Supervised:
  - for boring things that you don't want to do by hand
- Unsupervised:
  - for boring things that you don't want to do by hand
  - for scientific analysis



What does it have to do with astronomy?

Spectroscopic classes

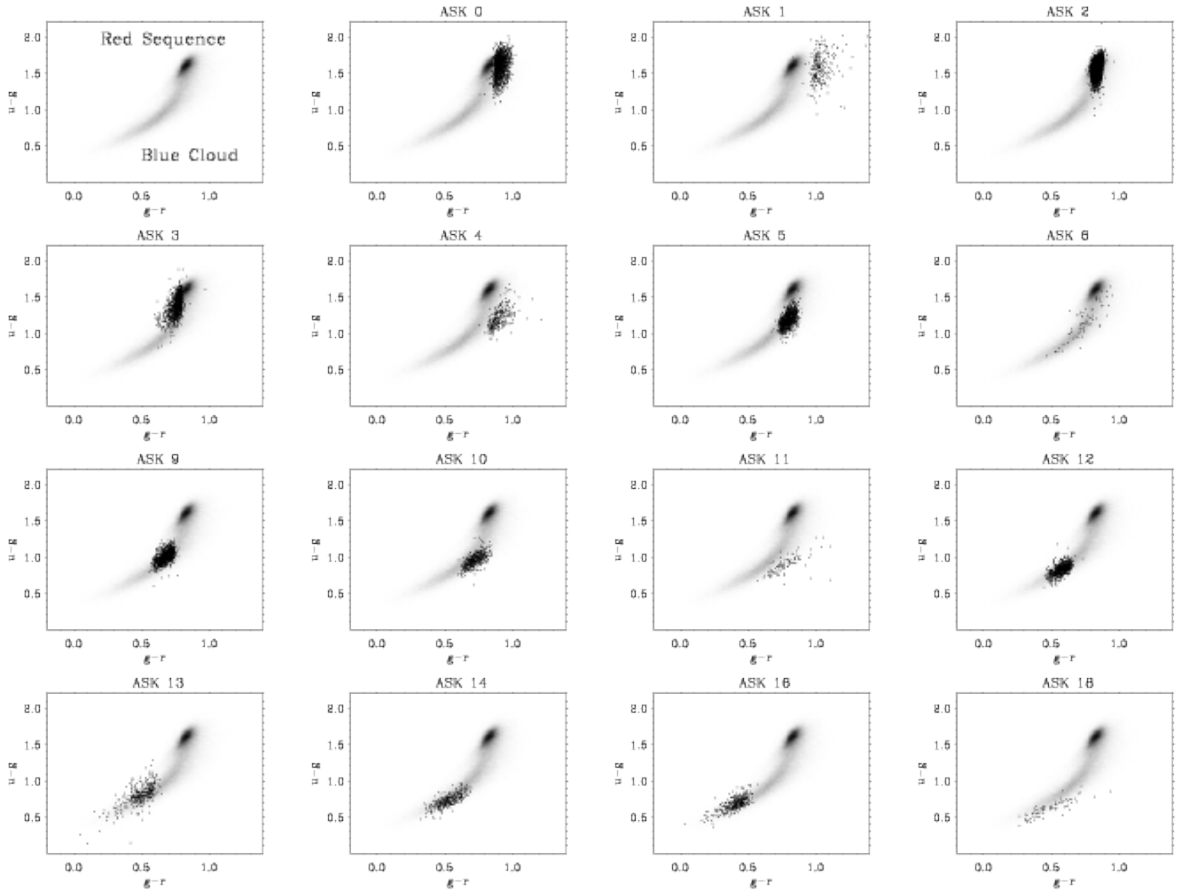


Automatic Spectroscopic K-means (ASK)

Definition of ASK classes

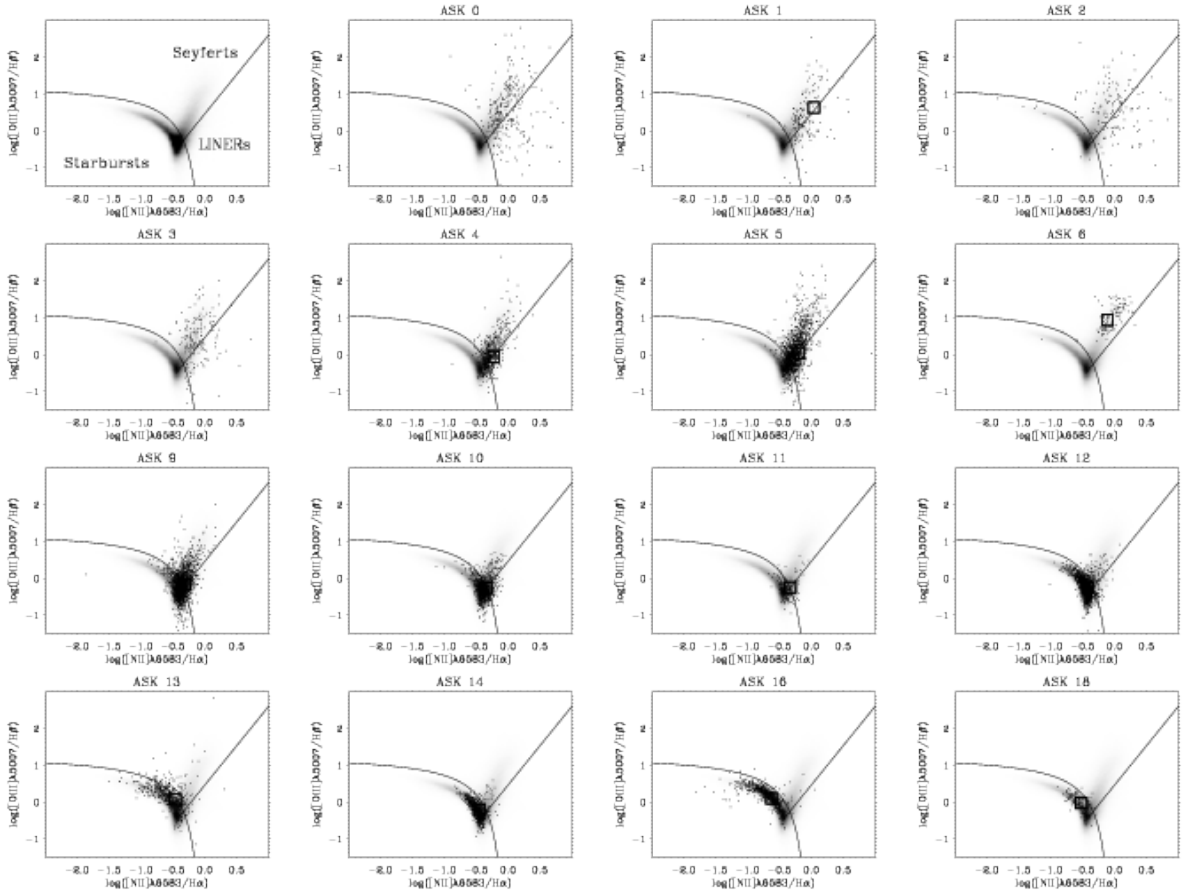
1. SDSS galaxy spectra
2. Rest-frame corrected
3. Normalized to g band
4. Feature selection
5. k-means

Automatic Spectroscopic K-means (ASK)



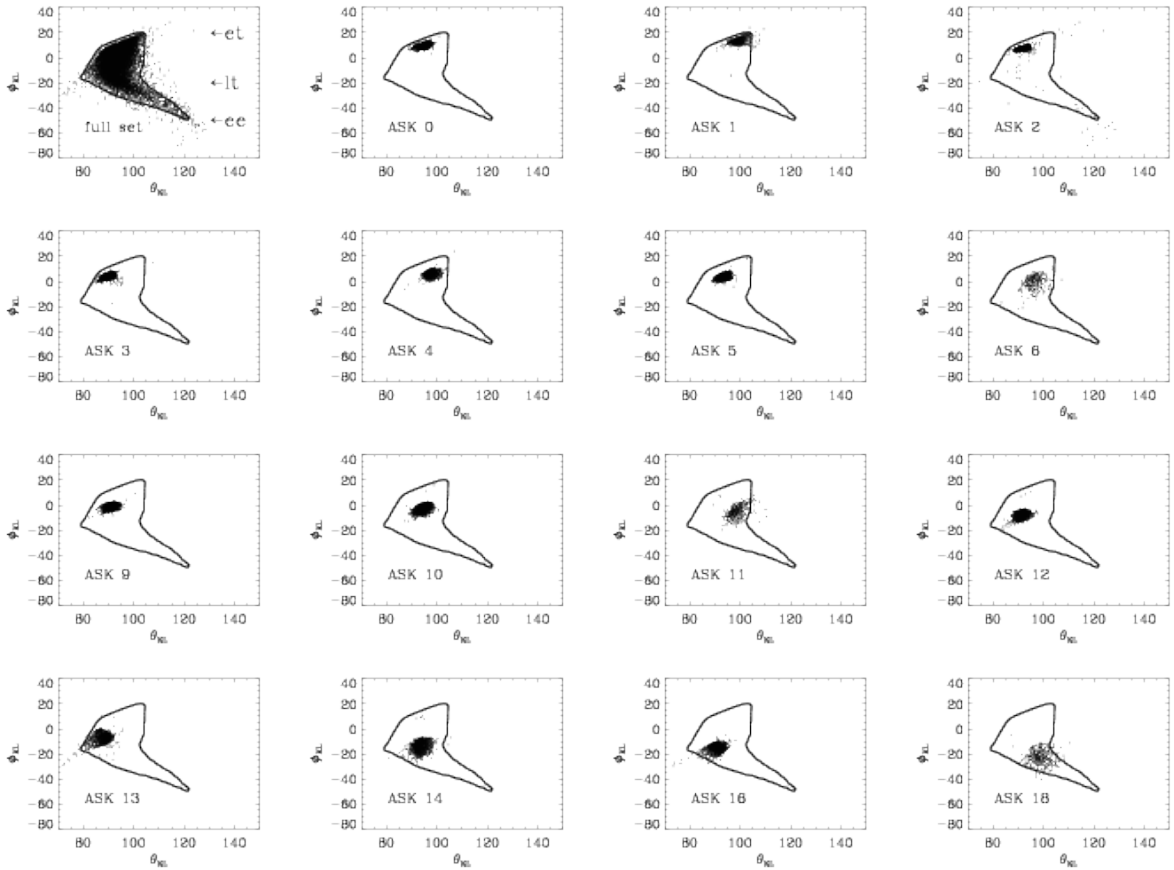
Sánchez-Almeida et al. (2010)

Automatic Spectroscopic K-means (ASK)



Sánchez-Almeida et al. (2010)

Automatic Spectroscopic K-means (ASK)

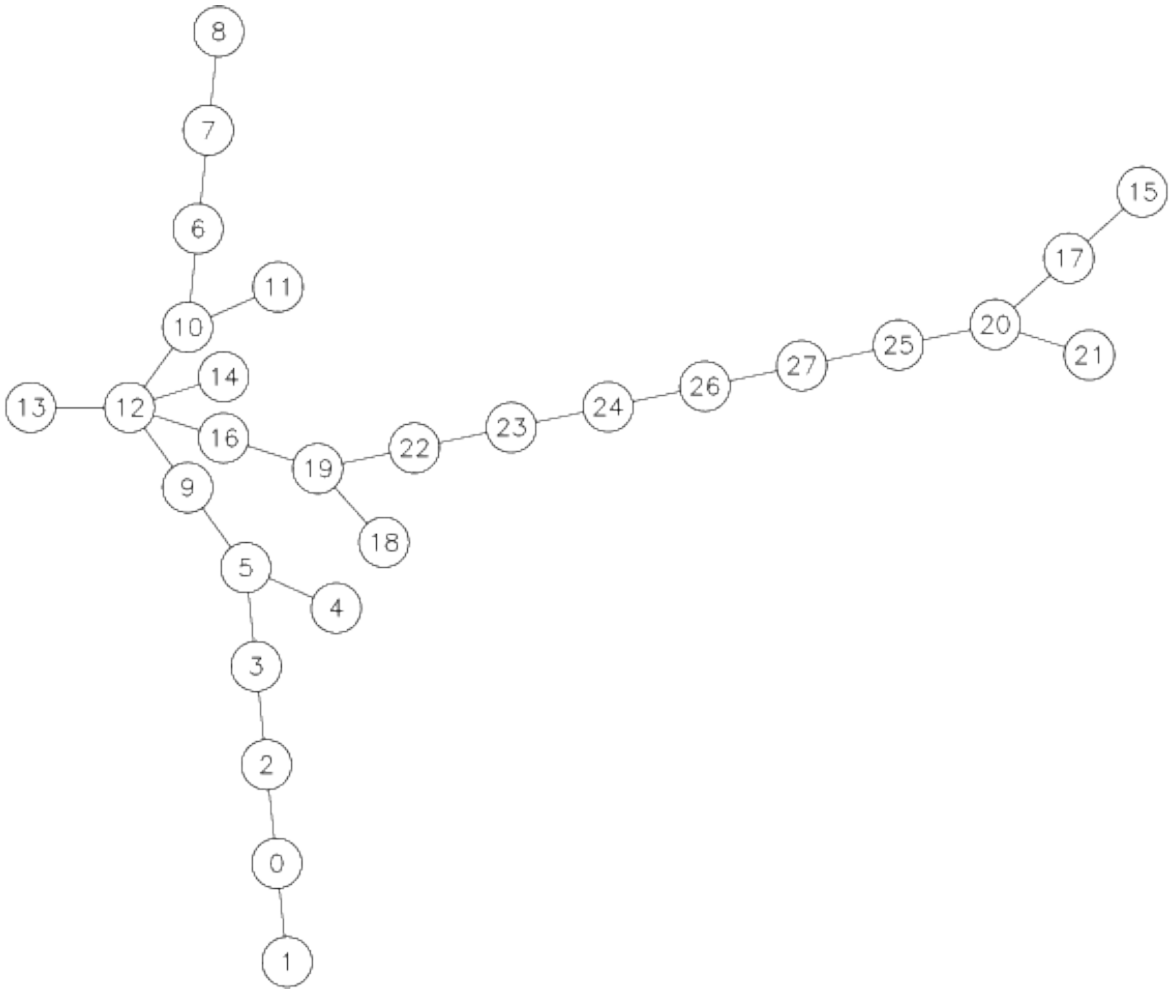


Sánchez-Almeida et al. (2010)

Distribution in spectral space

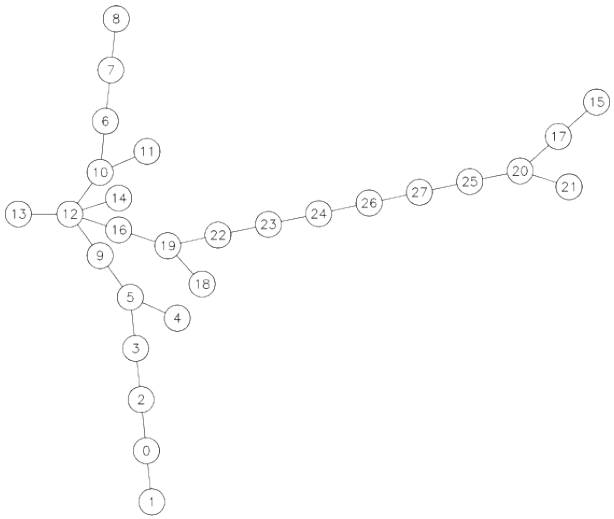
What do you think?

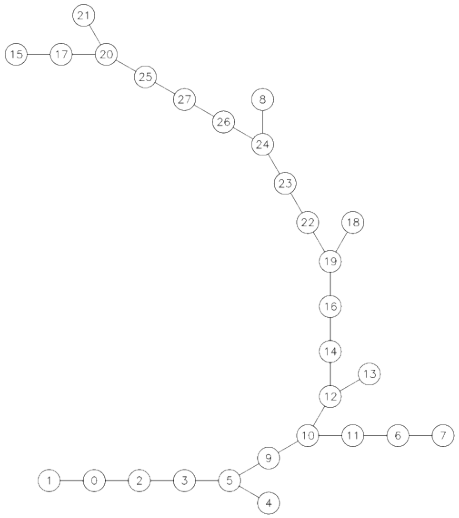
The MST of ASK classes

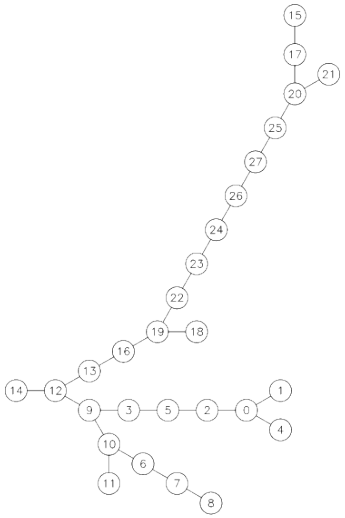


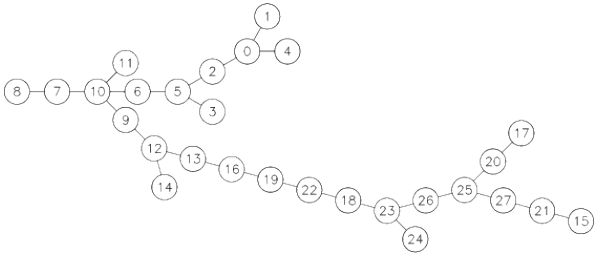
Ascasibar & Sánchez-Almeida (2011)

The MST of ASK classes



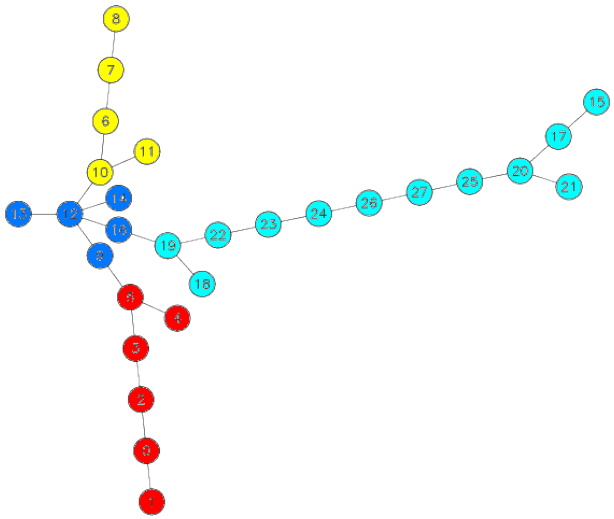


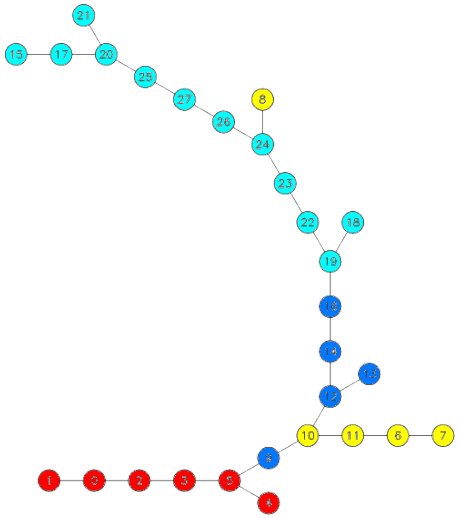


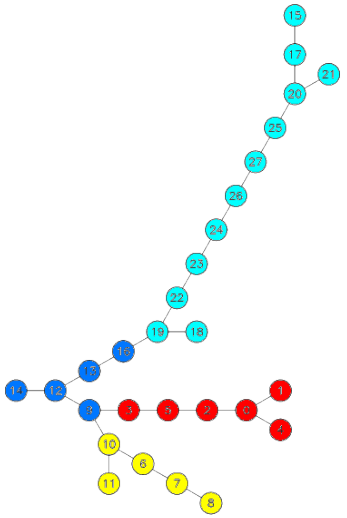


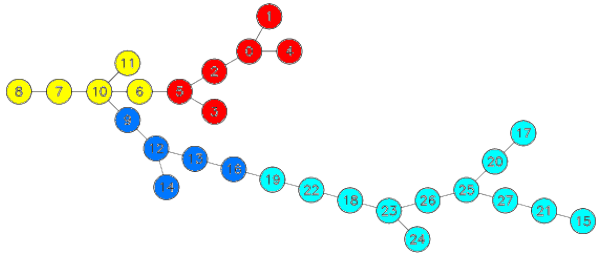
Ascasibar & Sánchez-Almeida (2011)

The MST of ASK classes



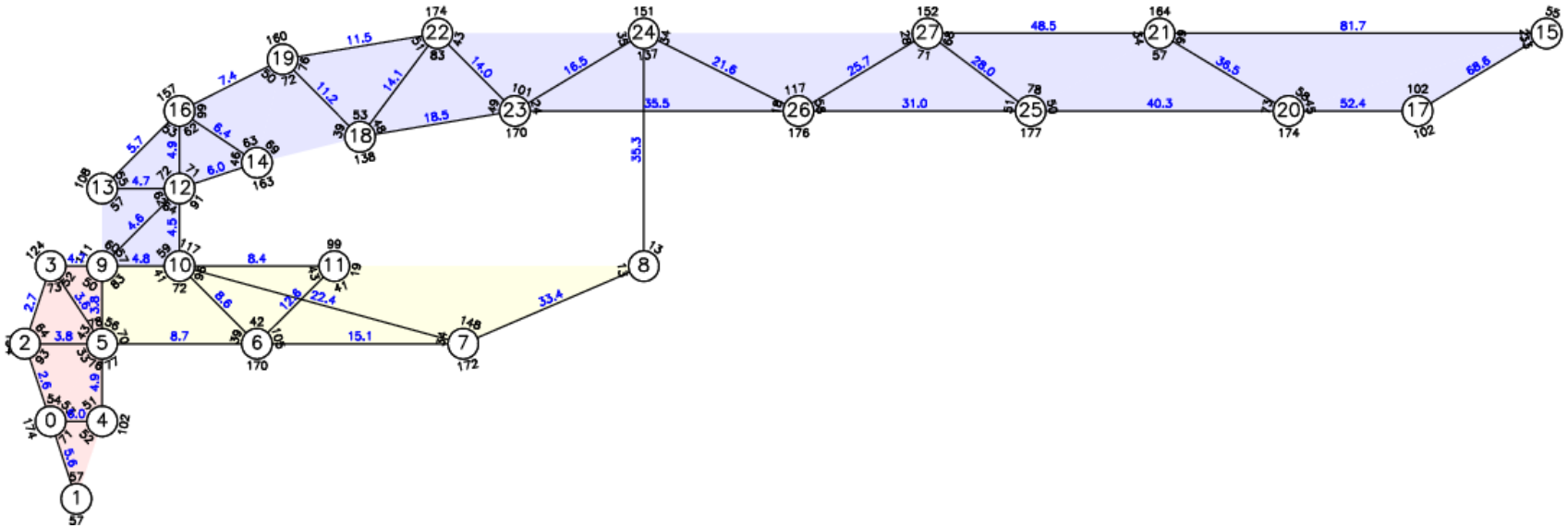






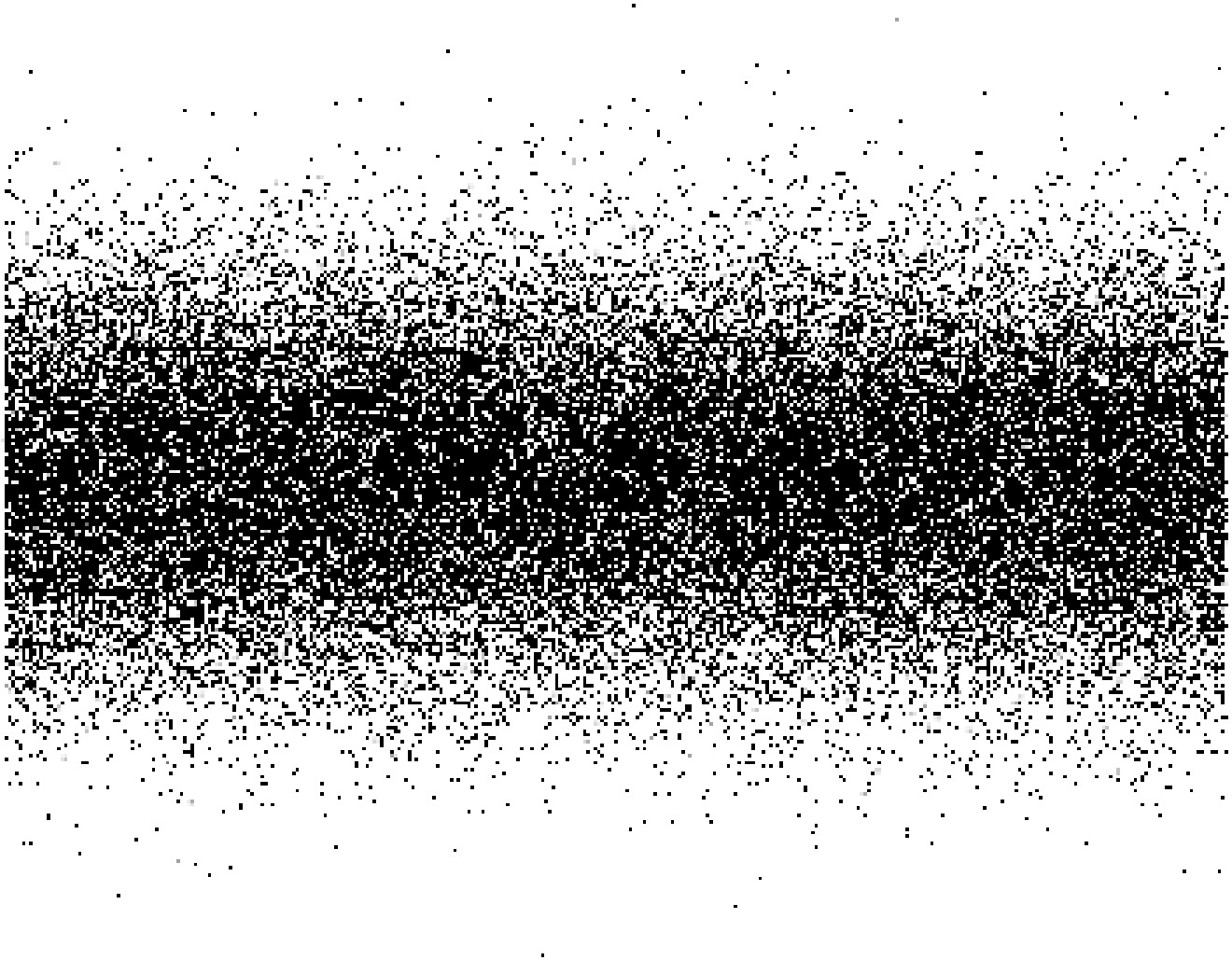
Ascasibar & Sánchez-Almeida (2011)

A spectroscopic sequence?

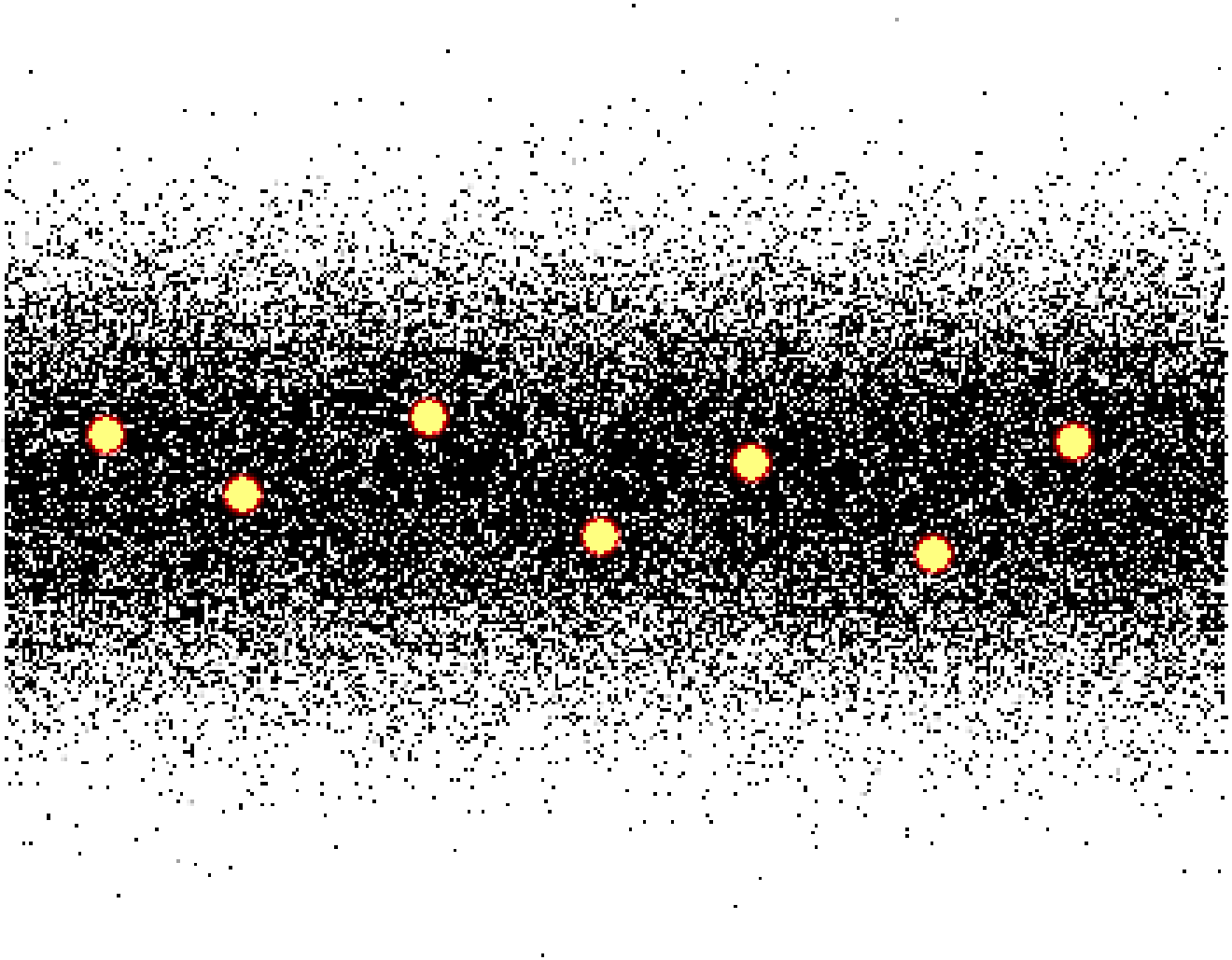


Ascasibar & Sánchez-Almeida (2011)

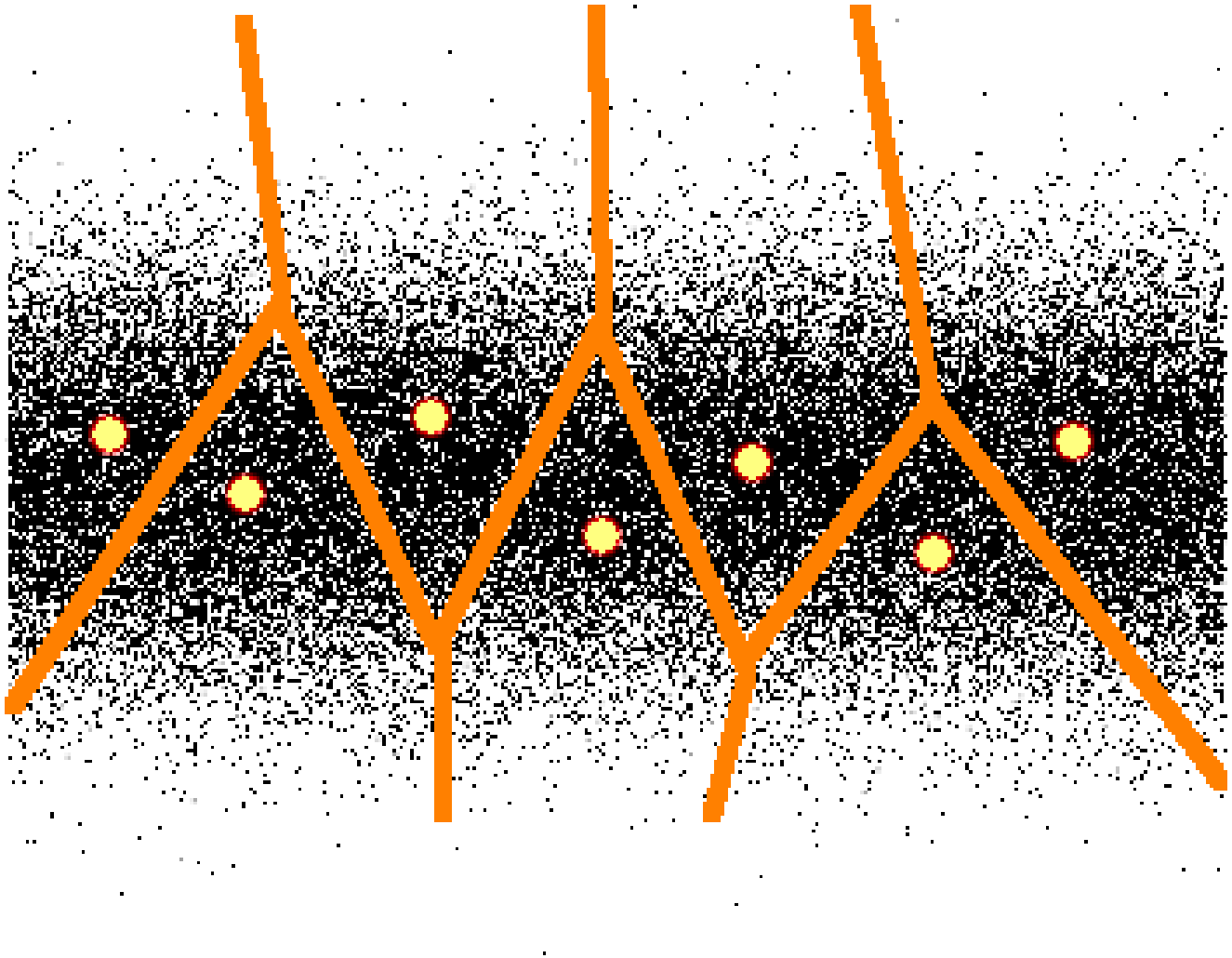
More on k-means



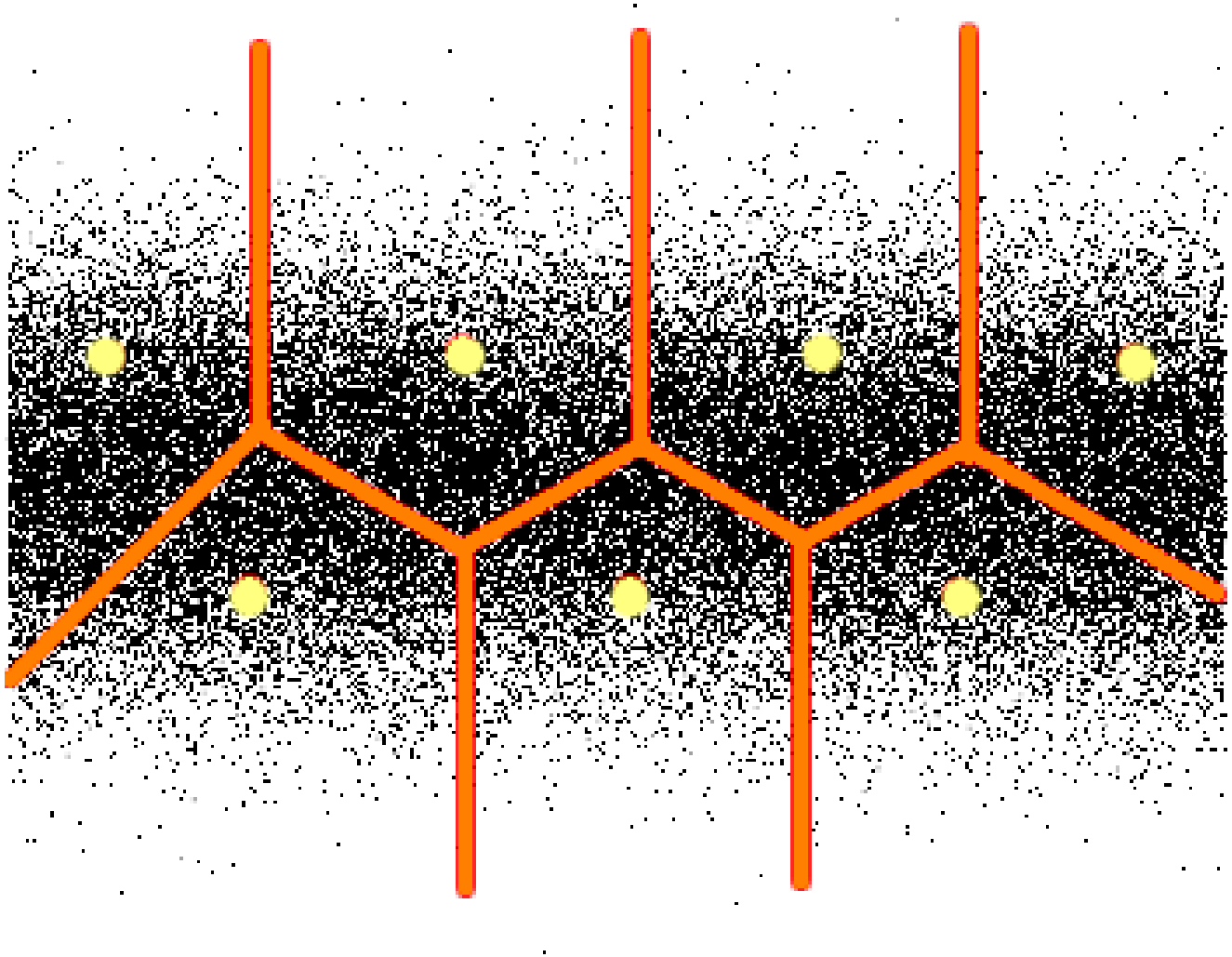
More on k-means



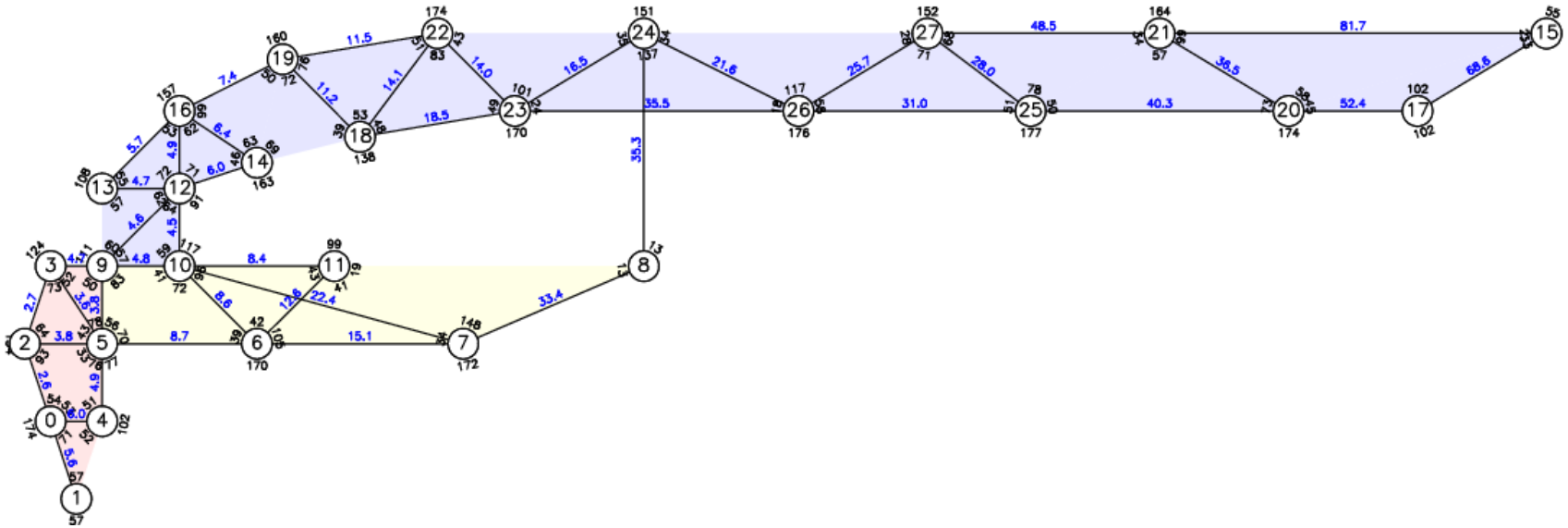
More on k-means



More on k-means



A spectroscopic sequence?



Ascasibar & Sánchez-Almeida (2011)

Physical interpretation

Normal galaxies

Normal galaxies form a well-defined, one-parameter sequence in spectral space

Active galaxies

Active galaxies form a separate branch that intersects the main sequence at the green valley

Phases of galactic evolution?

- Star formation
- Optical activity
- Red and dead

Summary

What to take home?

- There is a lot of **physical** information in the way data are distributed in multi-dimensional space
- There are several tools to retrieve it

What next?

- Learn more, e.g. http://scikit-learn.org/stable/user_guide.html
  http://www.astroml.org/sklearn_tutorial/
- Apply this approach to **your** problem