Data Mining in Astronomy

Final examination

Form: a 15-min presentation with a description of results, results (slides, pictures, catalog and a paper with the description of the work done) should be submitted before the presentation. During the presentation and in the written paper student should justify the selection of input data and software tools used in the study.

Subject: a description of the non-trivial scientific data processing based on the task performed by student. The choice of input catalogs, software tools and methods used in the study are up to the student.

Tasks:

1. Emission line classification of galaxies.

People classify galaxies into various types using emission lines and so-called diagnostic diagrams. The current classification scheme using optical emission lines is described in Kewley et al (2001, ApJS 132; see Eq 5-7), Kauffmann et al (2003, MNRAS 346) and Brinchmann et al (2004, MNRAS 351). This method is useful when you have a large range of emission lines available from [O II]3727Å to [S II]6716Å,6731Å. However at high redshift not all of these lines are available and no well-established techniques exist for classification of galaxies using only lines at wavelengths <5500Å.

The goal of this project is to carry out a systematic exploration of techniques to classify galaxies based on their emission lines and to devise a way to classify galaxies based only on the [O II]3727Å, Hbeta and [O III]4959Å,5007Å lines.

You are expected to **at least**:

- Define a sample of galaxies with emission lines – you are free to do this as you see fit.
- Define a training set classification
- Explore different classification techniques (more than 1!) - evaluate and compare strategies for classification.
- Create final tables with classifications and other information and draw E-R diagrams..
- Create a VO compatible table of the final dataset.

2. Quantify the ratio of supernovas of different types.

It is now broadly accepted that one type of Gamma-ray bursts (GRBs), the so-called long soft bursts, are the result of the core-collapse of some subset of massive stars (for an overview see Woosley & Bloom (2006, ARA&A 44, 507). These GRBs appear to be linked primarily to supernovae of the type Ib/Ic which are the explosive deaths of massive stars with no hydrogen envelope.

It is thought that the reason that these massive stars have no hydrogen envelope when they explode is because they have lost it as result of a stellar wind and that this depends on the star’s metal abundance.

The goal of this project is to gain some insight into this process by studying the fraction of supernovae of different types as a function of the metal abundance and absolute magnitude of their host galaxies. Ie. at the end we would like to see a plot that shows metal abundance or absolute magnitude on the x-axis and the ratio of the number of Ia supernovae in the bin to the total number of supernovae in that metal abundance/absolute magnitude bin, and so on for other types of supernovae (you might want to combine Ib and Ic).

For metal abundance the easiest is to use oxygen as a proxy for metal abundance and you can use the Pettini & Pagel (2004, MNRAS 348) calibration – their equation 1 or equation 3, but you are free to use other techniques if you can argue for them.

You are expected to **at least**:

- Define data sources for supernovae and host galaxies – make sure you choose data source for which you can calculate the quantities you need in the rest of the project!
- Calculate metal abundances and absolute magnitudes for the host galaxies.
- Plot the fraction of Supernovae of different types as a function of metal abundance, and magnitude of their host galaxies.
- Compare to previously published results – a good *starting* point here is Prieto, Stanek & Beacom (2008, ApJ 673).
- Create final tables with supernova classifications and information on host galaxies, and other information if you deem it necessary/useful and draw E-R diagrams.
- Create a VO compatible table of the final dataset.

3. Search for unresolved binaries in open clusters.

The task is to find a suspected unresolved binaries in open cluster using the systematic shift of the binary star position on CMD relative to the cluster's main sequence.

The task require creation of a compiled catalog containing photometry and membership information from literature. The catalog must include at least 5 bands and estimation of the membership of the star from literature. Student should plot CMD (based on the cross-identification in a number of catalogs) with clear main sequence of the cluster and using an estimation of the distance of each star from the main sequence to estimate the probability of this star to be an unresolved binary.

- Find an appropriate data sources for the cluster (photometric catalogs, catalogs with membership)
- Crossidentify input sources and create a compiled catalog
- identify main sequence part of the cluster isochrone and approximate it by the line in magnitude-color coordinates, compare this line with standard main sequence (Schmidt-Kaler Th., 1982, in Landolt-Bornstein VI)
- estimate probability of cluster members to be an unresolved binary by the distance from the main sequence to the star on N CMDs, N >2

Result: a catalog of stars of the open cluster with suspected binaries flagged (VOTable).

Objects: Pleiades, Hyades

4. Estimation of an age of open cluster

The task is to find an age of open cluster by comparison of the CMD of the cluster with theoretical isochrones. This task requires cross-identification of a number of photometric catalogs and using theoretical isochrones in the same photometric system.

- Select available sources with photometry and information about membership (if available)
- cross-identify catalogs and create a compiled catalog
- by comparison with synthetic isochrones (for example, Girardi et al., 2002, A&A 391, 195) find an estimation of an age for the cluster
- create VOTable with compiled catalog and VOTable with main parameters for the cluster (coordinates, identifiers, age, distance, E(B-V) estimation)

Result: a catalog of stars of open cluster (VOTable), CMD plot for the cluster and isochrones.

Objects: NGC 2243, Collinder 106, NGC 3324, NGC 6193, NGC 6231, NGC 2506, NGC 2420

5. Metal abundance gradients in galaxy disks

A consequence of the way the SDSS detect objects is that nearby galaxies are occasionally shredded into many pieces, each of which might be spectroscopically observed. This is a nuisance for statistical studies but it does mean that we get the chance to study the internal properties of these objects.

One topic in particular is the gradient of metal abundance with radius in galaxies. The goal of this project is to use the SDSS to measure how oxygen abundance varies with radius in nearby galaxies.

For metal abundance the easiest is to use oxygen as a proxy for metal abundance and you can use the Pettini & Pagel (2004, MNRAS 348) calibration – their equation 1 or equation 3, but you are free to use other techniques if you can argue for them.

You are expected to **at least**:

- Determine how oxygen abundance varies with radius in M101 which is the most spectacular example of shredding with many spectra across the disk.
- Find all nearby galaxies with multiple spectroscopic observations in the SDSS – where you are free to decide on a catalogue of nearby galaxies, but you should in general take nearby to mean that they have a recession velocity <5,000-20,000 km/s.
- Calculate gradients for those galaxies that have more than two spectra and compare all your results to the literature where possible.
- Create final tables with abundances and other information and draw E-R diagrams.
- Create a VO compatible table of the final dataset.