Machine Learning Experiments on Multi-Modal Astronomical Data

Farukh Rustamov

Supervised by: RNDr. Petr Škoda, CSc
CVUT, FIT

February 17, 2025

1 Introduction

In modern astronomical research, a large fraction of data is acquired through the analysis of celestial spectra. Spectral methods enable us to determine the chemical composition of stars or galaxies, to measure their velocities, to estimate star formation rates, and more. Knowing these parameters, in turn, makes it possible to classify objects: whether they are stars, galaxies, or quasars.

Below, we successively address three fundamental concepts:

1. Spectra and methods of spectral analysis,
2. Redshift (its definition and connection to the expanding universe),
3. Star Formation Rate (SFR) and its use in astronomy.

Finally, we demonstrate how combining this information helps distinguish between different classes of astronomical objects (quasars, stars, galaxies) in large sky surveys.

2 Spectra and Spectral Analysis

2.1 Definition of a Spectrum

A spectrum in astronomy is the dependence of an object's emitted intensity on wavelength (or frequency). If one passes the light through a prism (or a diffraction grating in a spectrograph), "white" light can be dispersed into many wavelength components (from ultraviolet to infrared). In practice, astronomers use specialized spectrographs attached to telescopes.

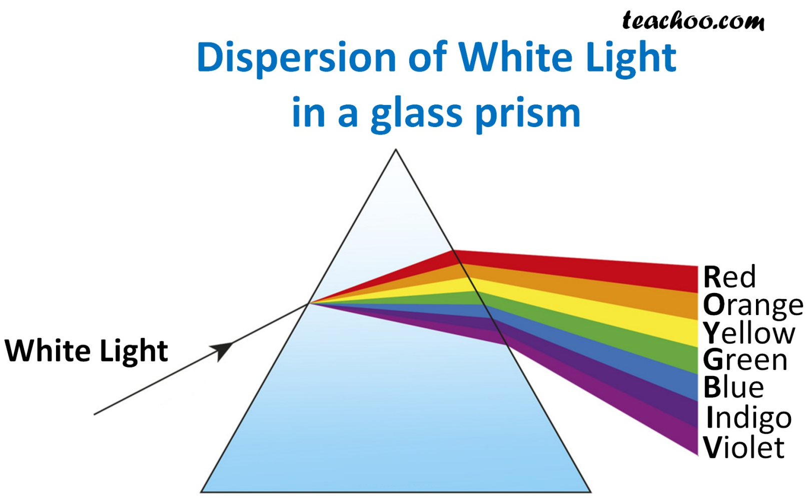

Figure 1: White-light decomposition through a prism.

2.2 Why Spectral Analysis Is Needed

- **Chemical Composition:** Every atom (hydrogen, helium, iron, etc.) has a characteristic set of spectral lines. By studying emission or absorption lines in the spectrum, one can "read off" an object's chemical makeup and derive certain physical parameters of the gas (temperature, density, ionization).

- **Velocity (Radial) Measurements:** If the lines are shifted from their "laboratory" wavelengths, it indicates the object's motion relative to us. Objects receding from us show redshifted lines, while those approaching us show blueshifted lines.

- **Other Parameters:** For instance, the star formation rate (SFR) can be derived from emission-line intensity (such as Hα) or from UV continuum emission, and so forth.

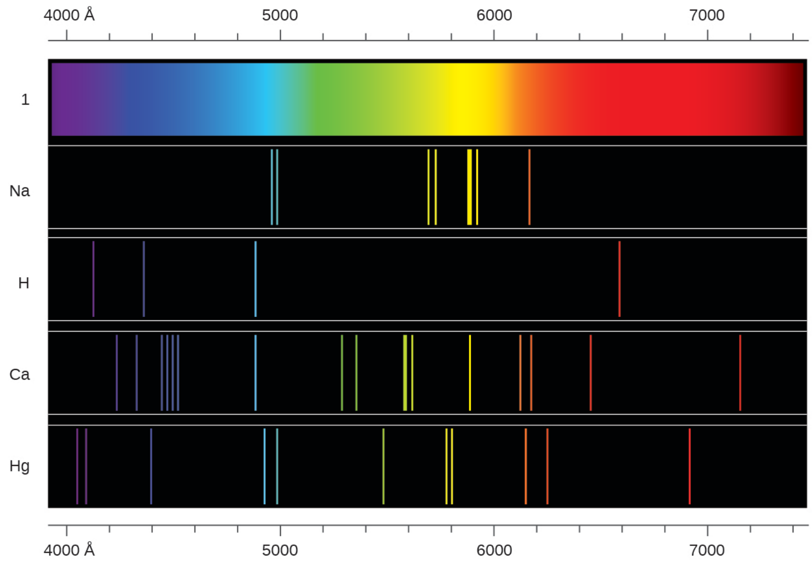

Figure 2: Example of atomic spectral lines for different elements.

For a more detailed overview of real spectroscopic applications, see the NASA reference on spectroscopy [1].

2.3 Typical Types of Spectra

- **Stellar Spectrum:** Often has strong absorption lines (e.g., Fraunhofer lines) corresponding to elements in the stellar atmosphere.

- **Nebula or Gas Cloud Spectrum:** Characterized by bright emission lines, which arise from ionized gas.

- **Galaxy Spectrum:** A "mixed" spectrum that combines the light of billions of stars, plus the emission of gas clouds, and sometimes an active nucleus.

3 The SDSS u,g,r,i,z Filters

One of the most influential surveys is the Sloan Digital Sky Survey (SDSS), which offers both photometric and spectroscopic data for millions of celestial objects. Its photometric system is built around five broadband filters: **u, g, r, i, z**. Each filter captures a different wavelength range:

- **u**: Near-ultraviolet (centered around 354 nm),

- **g**: Blue-green (centered around 477 nm),
- **r**: Red (centered around 623 nm),
- **i**: Near-infrared (centered around 762 nm),
- **z**: Infrared (centered around 913 nm).

From these five magnitudes (u, g, r, i, z), astronomers construct a *broadband* Spectral Energy Distribution (SED). By comparing an object's brightness in each filter, one can make an initial classification:

- **Stars** have typical color-color tracks distinct from galaxies,
- **Galaxies** show extended profiles in imaging and different color patterns,
- **Quasars** often exhibit unusually "blue" colors in certain redshift ranges or strong emission lines once spectra are obtained.

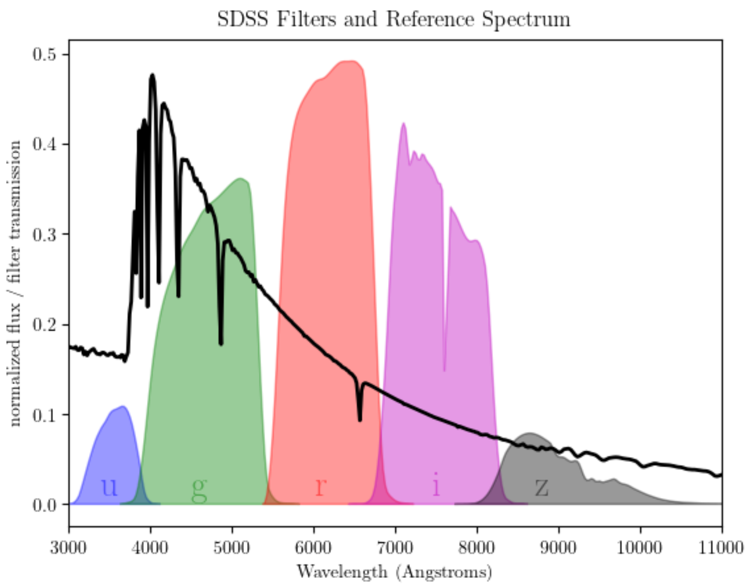

Figure 3: Transmission curves of the SDSS u,g,r,i,z filters.

Thus, the SDSS photometric system provides a powerful pre-selection of interesting sources, many of which are then targeted for follow-up spectroscopy to measure redshift and other parameters.

4 Redshift

4.1 Definition and Formula

Redshift (z) indicates how much the wavelength λ of emitted radiation is "stretched" if the source is moving away. Formally,

$$z = \frac{\lambda_{\text{observed}} - \lambda_{\text{rest}}}{\lambda_{\text{rest}}}, \tag{1}$$

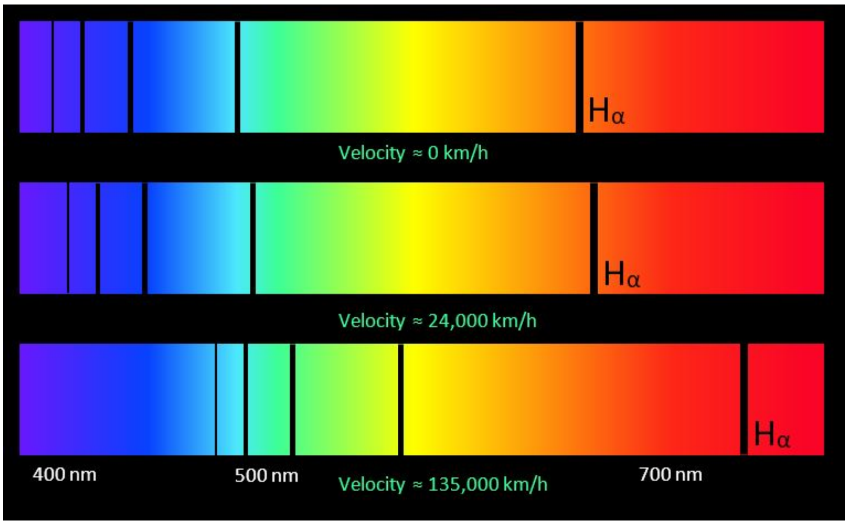

Figure 4: A schematic of redshift: an emission line at rest wavelength observed at a longer wavelength.

where

- λ_{rest} is the "laboratory" (reference) wavelength of a spectral line,
- $\lambda_{\text{observed}}$ is the measured (observed) wavelength.

If $z > 0$, the line is shifted toward red; if $z < 0$, it is blueshifted (the source is approaching).

4.2 Physical Nature of Redshift

Doppler Shift: If a source moves at speed v relative to the observer, photons can experience a change in wavelength (classical Doppler effect). For details, see [2].

Cosmological Expansion: For distant galaxies, the dominant effect is the expansion of space itself. The farther an object is, the greater its redshift (Hubble's law).

Practically, measuring redshift helps estimate how far away galaxies are (when z is moderate or large) and thus map the cosmic distribution of matter.

4.3 Why Mention Redshift in a Thesis

- **Galaxy Classification:** Knowing z clarifies how distant the galaxy is and how we see it in cosmic history.
- **Distinguishing Quasars:** Quasars often have large redshifts (up to a few), unlike typical galaxies with lower z.

5 Star Formation Rate (SFR)

5.1 What Is SFR

The Star Formation Rate (SFR) quantifies how many stars (by mass) form in a galaxy per unit time, typically expressed in $M_\odot\,\text{yr}^{-1}$.

5.2 How SFR Is Determined

A common method is to measure Hα emission in a galaxy's spectrum, since this line traces ionized hydrogen. More star formation implies more massive O-B stars, which ionize their surroundings:

$$\mathrm{SFR} \propto L(\mathrm{H}\alpha), \tag{2}$$

where $L(\mathrm{H}\alpha)$ is the total Hα luminosity. A standard calibration is

$$\mathrm{SFR}(M_\odot\,\mathrm{yr}^{-1}) \approx 7.9 \times 10^{-42}\, L(\mathrm{H}\alpha)\,(\mathrm{erg\,s}^{-1}).$$

UV or IR flux can also be used, depending on availability and dust content.

5.3 Why SFR Matters

- **Galaxy Classification:** Starburst or star-forming galaxies show high SFR, whereas elliptical galaxies are more quiescent.
- **Tracing Cosmic Evolution:** SFR helps reconstruct how stellar populations built up over cosmic time.

6 Combining Spectrum, Redshift, and SFR for Classification

Putting it all together:

- **Spectrum:** Informs us about the chemical composition (hydrogen, metals), presence of emission or absorption lines, overall continuum shape.
- **Redshift (z):** Indicates the distance or recessional velocity, helping discriminate nearby from distant objects.
- **SFR:** Shows how actively new stars form, i.e., whether a galaxy is star-forming, undergoing a burst, or is largely passive.

6.1 Example: Star vs. Galaxy vs. Quasar

- **Star:** Typically $z \approx 0$ (local). Spectrum features absorption lines characteristic of stellar atmospheres. SFR is not applicable to single stars.
- **Galaxy:** z can be small or large; spectrum includes combined stellar light plus potential emission from ionized gas. Hα strength can estimate SFR.
- **Quasar:** Often large z (e.g., 0.5, 1, 2, 3), with prominent, broad emission lines (Mg II, CIV) arising from accretion onto a supermassive black hole. SFR may be overshadowed by nuclear activity.

7 Conclusion

In a thesis, it is logical to explore:

1. **Spectral Analysis Basics:** How spectra are taken, why they are the main source of information about chemical composition and physical conditions.
2. **Redshift:** Definitions, why distant objects show $z > 0$, how expansion of space leads to large z for faraway galaxies.
3. **Star Formation Rate (SFR):** Connection to emission lines (e.g. Hα), ultraviolet, or infrared light, crucial for studying galaxy evolution.
4. **Object Classification:** Combining the shape and lines of the spectrum, redshift, and SFR to identify stars, galaxies, quasars.

This approach can span around nine pages of detailed explanation (excluding figures) and present a coherent overview of fundamental concepts: the spectrum, redshift, and SFR, as well as their usage in classifying cosmic objects. The SDSS u,g,r,i,z system is an essential building block in modern surveys, further connecting photometric properties with subsequent spectroscopic follow-up.

References

[1] NASA Article on Hubble Spectroscopy

[2] Example: Doppler Effect Reference on ScienceDirect (2018)