SDSS Celestial Objects Classification

Andrea Sessa Mat.850082

June 23, 2016

Abstract

Given the high dimensionality of the data provided by the Spectra instrumentation; main goal of the project is to provide a classfication of the celestial object observed by Spectra focusing on the data visualization and features selection phase.
Various different feautures selection tecnique are evaluated(Principal Component Anaysis(PCA), Kernel PCA, forward features selection, backward features selection). The different approaches are compared over a 'state-of-the-art' classifier(Support Vector Machine)
Final results shows that (TODO)

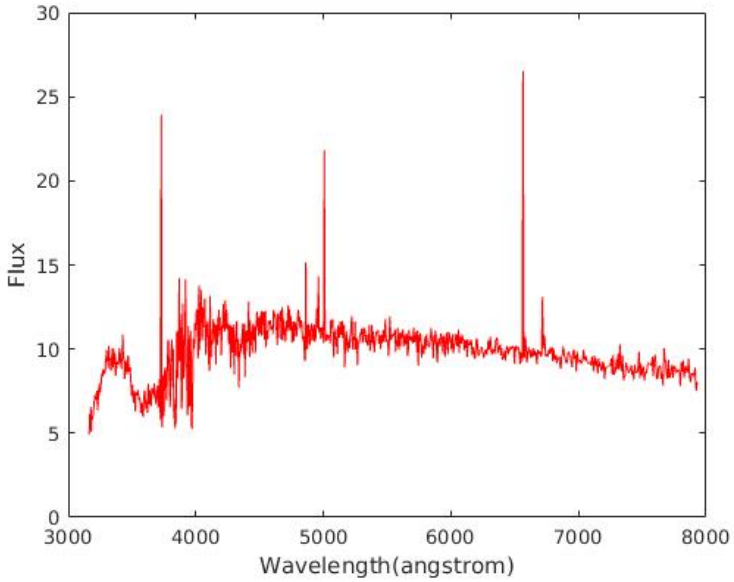
Contents

List of Figures

1 Introduction

Modern astronomy is concerned with the study of very distant celestial objects ie quasars, galaxies, stars, etc.
Often this type of classfication is performed by analyzing the spectrum emitted by such objects.
In general the emission spectrum of a chemical element or of a chemical compound is defined as the eletromagnetic radiation emitted when an atom or a molecule, of the object that we are observing, perform a transition from an high energy states to a low energy state. During the decadiment the atom or molecule the elettromagnetic is iradiated under the form of photon, the associated photon energy (also called flux) is proportional is equal to the energy difference between the two energy states involved in the decadiment.
The important element is that for a given atom there are many possible electron transition, and each of these transition has a specific flux associated.
The Sloan Digital Sky Survey(henceforth referred as SDSS) is a major imaging and spectroscopic survey using a dedicated 2.5-m wide-angle optical telescope at Apache Point Observatory in New Mexico, United States.
The collection of data started in 2000 and continues up to nowdays(latest data realease in June 2013). The dataset comprises almost 2 millions of spectra coming from diffrent objects.
Machine learning plays an important role in the task of classifying these objects, given the fact that many times samples include more than 1000 features.
Features in general includes information about the flux associated with a specific wavelenght.

Figure 1: Typical emission spectrum for a galaxy



The problem of feautures selection become very important in this scenario; this project will try to cast some light on the problem by evaluating different

approaches over a reducted version of the SDSS dataset.

2 Problem Formulation

2.1 Dataset

Write something about the dataset, number of samples, number of features, labels, in source

3 Methodology

4 Experiments

5 Conclusions

References

[1] Wikipedia, *Emission Spectrum*

[2] Wikipedia, *The Sloan Digital Sky Survey*