Estimating the Photometric Redshifts of Galaxies Using Regression Techniques

The Sloan Digital Sky Survey has been working for more than 20 years to make a map of the Universe. Measurements of large-scale structure in SDSS maps of galaxies, quasars, and intergalactic gas have had a huge impact for tests of the standard cosmological model that describes our understanding of the history and future of the Universe.

Let’s have a review over the details about many of the topics SDSS-V will address.

Galactic archelogy

Gas in the galaxy: The energy exchange between interstellar gas and stars that occurs as stars form reflects the detailed physics of gas accretion and stellar feedback.

Star forming regions:  classify and characterize roughly 100,000 young stars in the Galaxy.

Multi-star and planetary systems: SDSS-V will measure the environmental dependence of the multi-star fraction in the disk and stellar clusters.

White dwarfs: SDSS-V will identify and characterize.

Measuring Redshift from Spectroscopy

We all know that Hydrogen is the most abundant element in the universe, and it is often seen in galaxies. This spectrum comes from a galaxy, it shows strong emission lines. The hydrogen lines are already identified for you: the tallest peak is the alpha line, and the tall peak to its left is the beta line. The spectrum of such a region shows a pattern called the Balmer series of lines in emission. Now we should measure the shift of any one of those lines with respect to its expected wavelength, as measured in a laboratory on Earth.

Measuring Redshift from Photometry

We will be using flux magnitudes from the Sloan Digital Sky Survey (SDSS) catalogue to create color indices. Flux magnitudes are the total flux (or light) received in five frequency bands (u, g, r, i and z). The astronomical color (or color index) is the difference between the magnitudes of two filters, i.e. u - g or i - z. This index is one way to characterize the colors of galaxies. For example, if the u-g index is high then the object is brighter in ultra violet frequencies than it is in visible green frequencies. Color indices act as an approximation for the spectrum of the object and are useful for classifying stars into different types. Here F(\lambda) is the filter transmission, and S(\lambda) is the flux at wavelength \lambda. The constant m_{ref} encodes the calibration of the telescope.

Astronomers generally work in terms of the *color*, defined as the difference of magnitudes between two different filter bands. This is because the constant m_{ref} can be difficult to calibrate from telescope to telescope or from night to night. Subtracting two magnitudes reduces this uncertainty.

Machine learning

Sloan Digital Sky Survey that observed millions of galaxies. It's relatively straightforward for a person with a help of a computer to measure a red shift from a galaxy with an observed spectrum as we saw earlier. But many galaxies have not been observed spectroscopically, we only have images. In addition, the large number of galaxies in these surveys makes this task impractical to do by hand.

we're going to use machine learning to calculate the red shift of galaxies from their measured colors. This task is ideal for machine learning and there has been a substantial amount of research in this area.

Decision Tree

Each node represents a decision that the robot needs to make (or assess) to reach a final decision. In this example, the decision tree will be passed a set of input features (Outlook, Humidity and Wind) and will return an output of whether to play or not.

Random Forset

Basically combination of several dt.

Geometrical view

For a more precise look, split to particular areas and allocate the mean amount to those data.

Contour map of the redshifts

It shows that we get reasonably well-defined regions where redshifts are similar. If we were to make a contour map of the redshifts in the color index vs color index space, we would be able to get an estimate of the redshift for new data points based on a combination of their color indices.

Decision Tree

We can see how our calculated colour indices are input as features at the top and through a series of decision nodes a target redshift value is reached and output.

Optimizing process for Decision Tree

In order to see how the tree is overfitting we would like to examine how our decision tree performs for different tree depths. Specifically, we would like to see how it performs on test data compared to the data that was used to train it.

However, as the model overfits we see a difference in its accuracy on the training data and the more general testing data.

Final Result of Decision Tree Algorithm

we can compare the predictions their corresponding target values later when calculating the median difference and plotting the predicted values against actual values.