Comparing photometric redshift to other methods of measuring extragalactic distances

IB Physics Extended Essay

Toshinari Tong Chun Shing

Contents

Introduction

From ancient times, human have looked up the skies and wondered if one day we can reach the stars. As technology improved, we knew more and more about our universe, but the urge to one day travel to the vast universe never ceased. Finally, we managed to reach one of the many objects in the night sky: the moon in 1969. We still have this desire today, leading us to land rovers on Mars and develop commercial spaceflight.

However, modern astronomers are motivated by another reason to measure cosmic distances: to understand the fundamental properties of the universe. In 1929, Edwin Hubble discovered the universe is expanding by observing Cepheid variables, leading to the Hubble constant. By measuring vast extragalactic distances more and more accurately, astronomers are striving for more accurate value of the Hubble constant. This allows us to discover more about the past, present and future of the whole universe.

Finally, machine learning is a rapidly growing field because of the evergrowing amount of data and the evergrowing need of efficiently processing such data. In astronomy, data collected from surveys and telescopes has increased rapidly in the past few decades, and astronomers have set their eyes on machine learning to drive their data-driven quest on understanding the universe.

In this essay, I will train a neural network that performs photometric redshift, and compare it with other methods of measuring vast distances, including the period-luminosity relation of Cepheid variable stars, and the magnitude of type Ia supernovae.

Background

Firstly, let's introduce the magnitude system, which is a logarithmic measure of brightness. This is convenient because human perceive brightness logarithmically. There are two kinds of magnitudes used for stars and galaxies: the apparent magnitude m is the brightness as it appears in the sky as observed on Earth, while the absolute magnitude M is the intrinsic brightness of an object, defined as the brightness of the object if it was placed 10 parsecs away from Earth. The magnitude system is defined as follows: an increase in 5 magnitudes correspond to 100 times the brightness (known as Pogson scale, defined by Norman Pogson in 1856).

$$m_1 - m_2 = -\sqrt[5]{100}\log_{10}\frac{L_1}{L_2} \approx -2.512\log_{10}\frac{L_1}{L_2}$$

where m_i is the magnitude, and L_i are the luminosity of the object.

Traditionally, the zero-point of the magnitude scale is calibrated as the brightness of the star Vega, but other systems may have different zero-points. Moreover, the scale works 'in reverse', so brighter objects have a smaller magnitude. As the brightness scale has no upper or lower limit, negative magnitude are also possible, such as Sirius, which has an apparent magnitude of -1.46. Finally, if there is a subscript, for example M_v, that indicates the absolute magnitude measure in the visual V broadband filter.

Since stars are point sources in the sky, the magnitudes are the brightness for that 'point'. However, galaxies are elongated objects, so magnitudes for galaxies are obtained by using its integrated brightness, or 'adding up' the total brightness in the area it covers. (further explored later).

Another concept that has to be introduced is cosmological redshift. This is a measure of distance that deals with intergalactic distances involving galaxies, supernovae and quasars. Cosmological redshift is the phenomenon in which light emitted from extremely faraway objects are stretched while travelling through the expanding universe, like a drawn arrow on a stretching balloon. Light shifts to the red side of the spectrum as it is stretched, hence the name. The cosmological redshift z is defined mathematically as:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}}$$

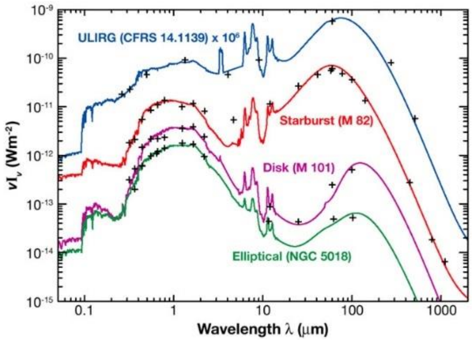
where λ_{obs} is the wavelength of light received by the observer and λ_{em} is the wavelength of light emitted by the source.

This is not to be confused with the Doppler redshift, which is more akin to Doppler shift of sound waves. The Doppler redshift happens as objects moving in extremely high speed emit light, shifting it red if it is moving away from us and shifting the light blue if it is moving towards us. On the other hand, cosmological redshift occurs because space itself expands, stretching the light in the process. From this point onwards 'redshift' will refer to cosmological redshift.

Photometric redshift

Spectroscopy

Before discussing photometric redshift, redshift obtained from spectroscopy (spectroscopic redshift) must be addressed as both methods are similar and spectroscopic redshift is also needed for photometric redshift. Spectroscopy is the study of electromagnetic radiation in terms of its constituent wavelengths. In a spectrograph, light is passed through a slit mask and dispersed by diffraction gratings, obtaining the spectral energy distribution (SED) of celestial objects. SEDs of 4 different types of galaxies are shown in the diagram to the right, each with different unique features.



https://ui.adsabs.harvard.edu/abs/2004PhDT...
......8G/abstract

The wavelength (x-axis) is usually measured in Angstrom (Å), where 1 Angstrom is equal to 10^{-10} meters. Radiant flux density (y-axis), referred to as flux, is the total amount of energy that crosses a unit area per unit time, usually Watts per meter squared. In this essay, data from SDSS is used, in which flux is measured in nanomaggies. A star of 1 nanomaggie has an apparent brightness magnitude of 22.5, and is related to the apparent brightness as follows:

$$m = 22.5 - 2.5\log_{10} f$$

where m is the apparent brightness and f is flux in nanomaggies.

Redshift (z) is the elongation of the wavelength of electromagnetic radiation of a faraway object due to the expansion of universe as mentioned previously. Therefore, the measured SED is shifted to the right compared to the actual SED[1]:

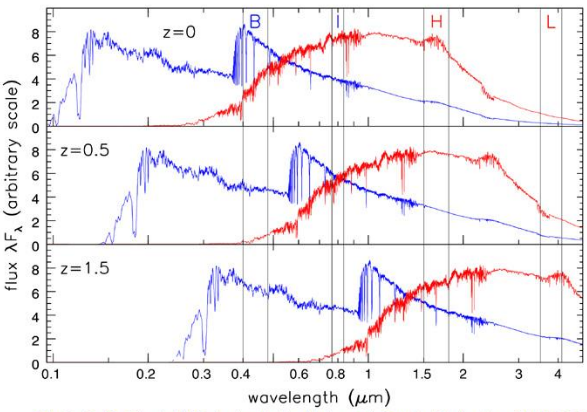


Fig 8.12 (S. Charlot) 'Galaxies in the Universe' Sparke/Gallagher CUP 2007

To find out how much the spectrum is shifted, astronomers look for characteristic peaks and dips corresponding to emission or absorption lines in the spectrum. Once the

[1] Image from *http://www.astro.wisc.edu/~sparke/book/ch8figs/ch8_index.html*

corresponding features in the redshifted spectrum is identified, the redshift can be easily calculated from the previously mentioned formula $z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}}$.

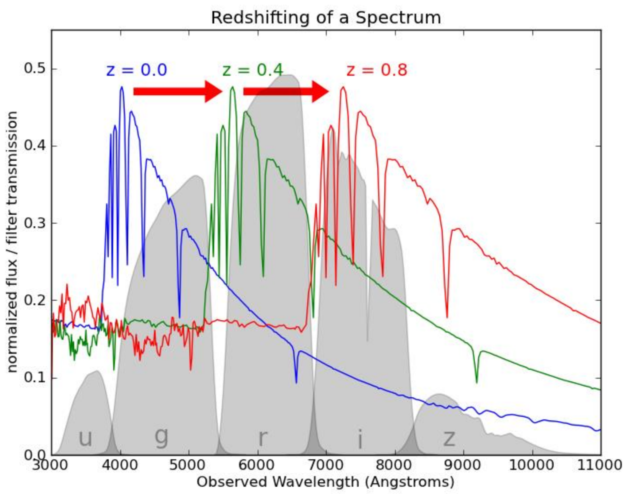
In galaxies, emission lines are often found in star-forming regions, where photons were emitted due to the excited atoms going from a high energy state to a low energy state. Each element has different energy levels, which emit photons of different wavelength ($E = hf$), so they can be used as characteristic peaks. On the other hand, absorption lines (dips in the SED) are formed when radiation from the center of a galaxy is absorbed by the interstellar medium or gas clouds.

Photometry

Spectroscopic redshifts of individual galaxies are important, but it is severely insufficient for modern astronomy because spectroscopy is a very time-consuming and expensive process as the spectrograph must be precisely positioned to obtain a SED in 1 single point in the sky. In order to produce redshifts for ideally large amount of objects more efficiently in our universe, the concept of photometric redshifts was born.

In photometry, the flux from large chunks of the sky is captured at once, making it much more efficient than spectroscopy. CCD (charge-coupled devices) cameras are used, which is essentially a grid of CCD photometers (devices that count incoming photons). Typically, different light filters are used in combination to provide more information. In SDSS, 5 broadband filters u', g', r', i', z' are used, capturing the total flux in different sections of the SED, producing 5 images for each chunk of the sky. There are also other filter systems, such as the traditional UBV system or that shown in the previous diagram (B,I,H,L).

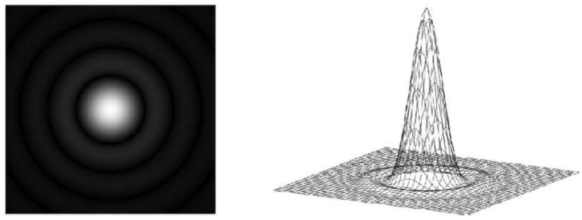
The 5 numbers of flux can be thought of as a SED with extremely low resolution, and the method of photometric redshift is trying to estimate the redshift from these values.



When photometric redshift were first investigated, statistical methods such as template fitting and quadratic regression were used to empirically find the redshift. Nowadays, astronomers are using the new technique of machine learning to effectively let the machine 'learn' the pattern between the input (flux from broadband filters) and the output (redshift), which is also efficient for processing large amount of data.

Magnitude fits

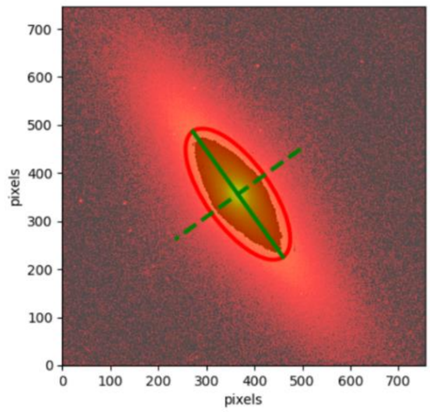
The magnitude (previously mentioned integrated brightness) of a celestial object is computed by adding up the photon counts in the region it covers. For stars, this is easy because they can be treated as a point, and the surrounding region of the point in the shape of an Airy disk can be summed to obtain the magnitude. (the z-axis represents photon count in the below diagram)



A B

However, the region covered by a galaxy is more complicated because it is elongated, either because the galaxy itself is an ellipsoid, or because the galaxy is disc-shaped and is inclined at an angle. To calculate the apparent magnitude, astronomers model a theoretical profile of the galaxy, then use the profile to calculate the magnitude. This way, light from neighbouring celestial objects can also be filtered out, making the result more accurate.

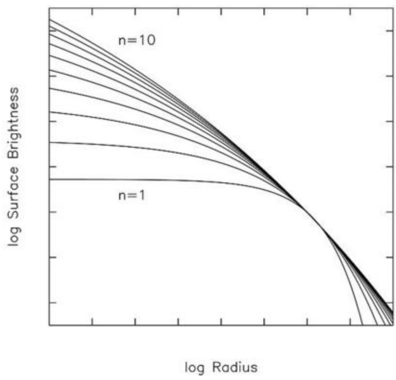
In the following section, the Python package `mgefit`[2] is used to fit the magnitude of galaxy NGC 4342 as an example. Firstly, the ellipse shape must be first identified on the photograph. The center is chosen among the brightest point in the photograph. Then, an algorithm identifies the angle, semi-major and semi-minor axis of the ellipse covered by the galaxy (the red and green lines). It also finds the isophote (a curve, or a contour line where the flux of all points on the curve are same) which contains half the total flux, which will be useful later. The isophote area is represented as the dark part inside the ellipse.



After that, the profile is modelled radially according to the Sersic profile:

$$I(R) = I_0 e^{-b_n \left(\frac{R}{R_e}\right)^{\frac{1}{n}}}$$

It is a function that describes how the intensity of a galaxy varies with distance R from the center. It can be thought of as the cross-sectional area of the angular 'slice' if the galaxy is a 'cake'. I_0 is the intensity at the center of the galaxy, and R_e is the radius of the isophote containing half the total flux. n is called the Sersic index, which controls the degree of curvature of the profile. The smaller the value n, the less centrally concentrated the profile is and more bent inwards the whole profile is[3]:



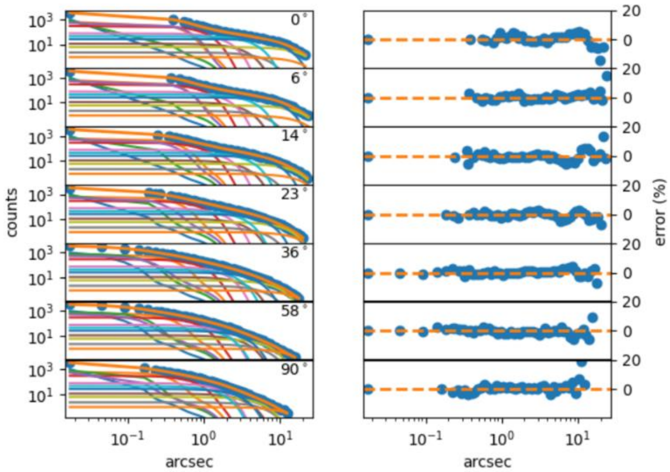
---

[2] from Cappelari 2002

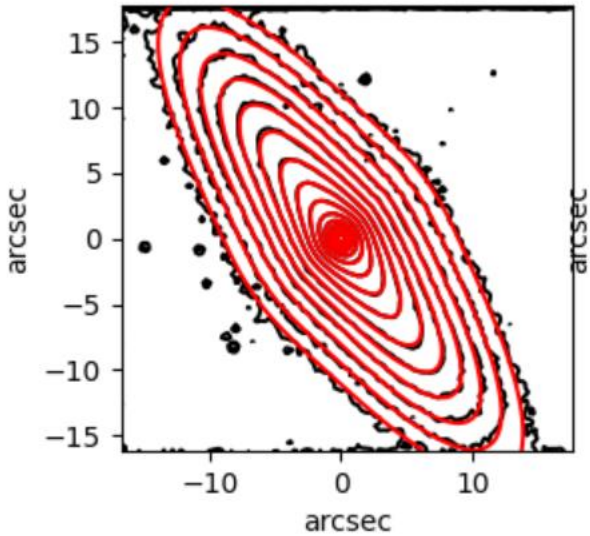
[3] from Burke 2013

$n = 4$ is an approximation for elliptical galaxies, called the de Vaucouleurs profile, while $n = 1$, or the exponential profile, is usually preferred for spiral galaxies. Sersic indices are in the range $0.5 < n < 10$ for most galaxies, and bigger and brighter galaxies tend to be fit with larger n.

`mgefit` fits these profiles in different angles of the ellipse to determine the best fit.:



Finally, the radial profile can be combined with the ellipse to produce a final magnitude fit, and the magnitude of the galaxy is calculated by integrating the fit. (In the diagram, the red lines are isophotes and the intensity is increasing as the isophote is closer to the center.)

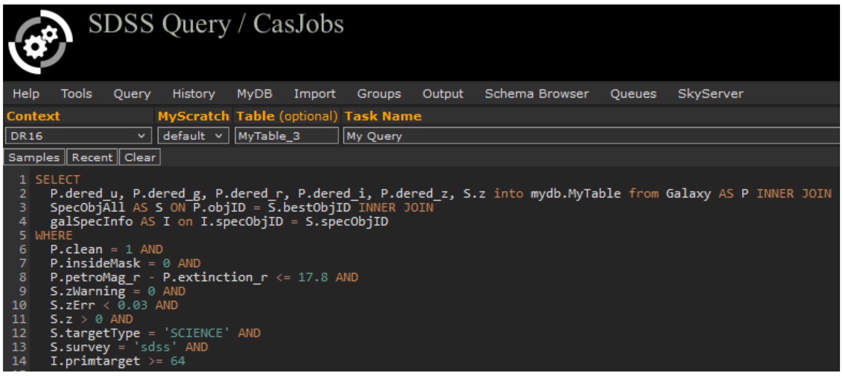


In SDSS, the above process is automatically applied in a pipeline by sophisticated computer systems. The magnitude that has a higher likelihood from de Vaucouleurs model and the exponential model is chosen in SDSS, and it is all stored in their database so it can be directly queried and downloaded.

Galaxy selection

The machine learning technique of neural network will be used as it can learn non-linear relationship between the input and output. However, a good dataset must first be chosen to train a neural network with good results. `dered_u`, `dered_g`, `dered_r`, `dered_i`, `dered_z` from the photometry database are chosen as inputs, which are the model fit magnitudes subtracted by extinction, which is the absorption or scattering of light by dust and gas in the interstellar medium between the celestial object observed and the observer. On the other hand, `z` from the spectroscopy database is chosen as the true output, as spectroscopic redshift is more accurate. The galaxies are chosen from SDSS DR16[4] (Data Release 16) based on the following criteria:

- the photometry and spectroscopy is clean and has no errors
- the dereddened Petrosian magnitude is ≤ 17.8 (reduces fluctuations of the dataset[5])
- the galaxy is observed by SDSS (other surveys are also included in the database)



Neural network

Generally, a machine learning method is first trained by a training dataset, where it tries to improve itself each iteration by comparing its result with the model output of the training dataset. After the learning process, it is evaluated by testing it on a completely different dataset, called the testing dataset, to actually grade if it learned the relationship between input and output or it just learned the training dataset. The downloaded dataset of 551617 galaxies are divided into the training set and test set by the ratio $3 : 1$. The training set is scaled and shifted such that its mean is 0 and its standard deviation is 1, as the learning model employed assumes so. The test set is transformed the same way, but its mean might not be 0.

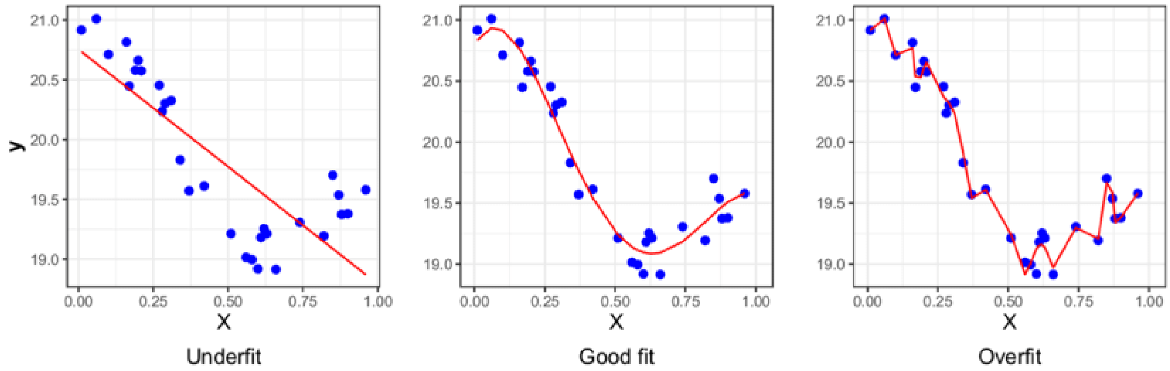
After trying different machine learning algorithms including decision trees and k-nearest neighbours, neural network was chosen for its superior performance. A neural network consists of an input layer, an output layer and 1 or more hidden layers. Each edge represents a weight, and a layer is transformed to another by summing up the value of node i in the previous layer times the weight of edge from i to the current node, similar to vector-matrix multiplication. The hidden layers allow the network to learn

[4] http://skyserver.sdss.org/CasJobs

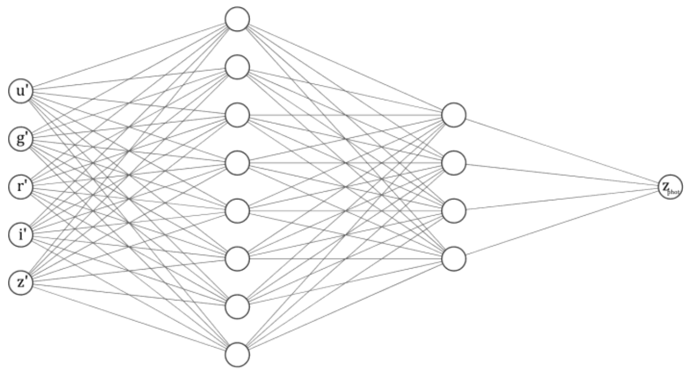
[5] Strauss, Michael A. et al 2002

non-linear relationships as it is a combination of multiple vector-matrix multiplication. The neural network tries to minimize the loss function, which measures how different its output and the model output is. In each iteration, all the weights are slightly changed according to the neural network to gradually minimize the loss.

A neural network also have a lot of hyperparameters to tune, which are parameters of the network itself, such as its learning rate, hidden layer sizes, activation function etc. If the hyperparameters are set incorrectly, the neural network may overfit, where it 'learns' too much about the training dataset and is unable to generalize to other data, or it may underfit, where it does not learn the relationship enough. I tried many different values and combinations of the hyperparameters to produce the best possible result.

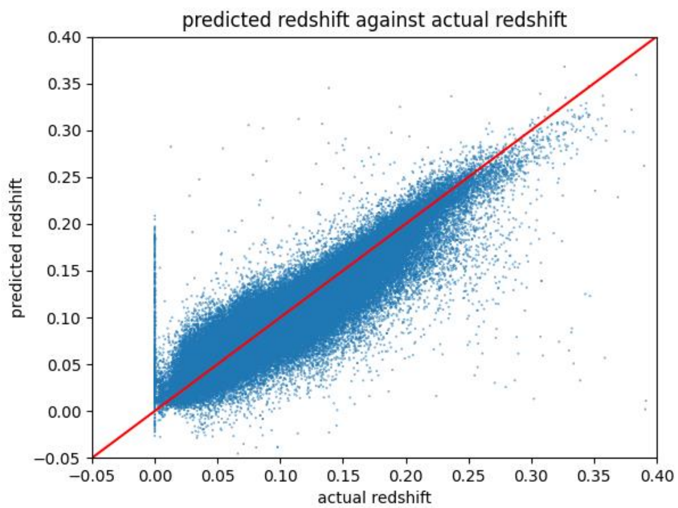


The resulting neural network is implemented by `scikit-learn`, a library of machine learning algorithms. It uses the loss function of mean squared error, the activation function of the hyperbolic tangent function, and has a structure as follows[6]:



The increased number of nodes in the first hidden layer allows the neural network to extrapolate relationships between the input, and the second hidden layer condenses the results into the redshift.

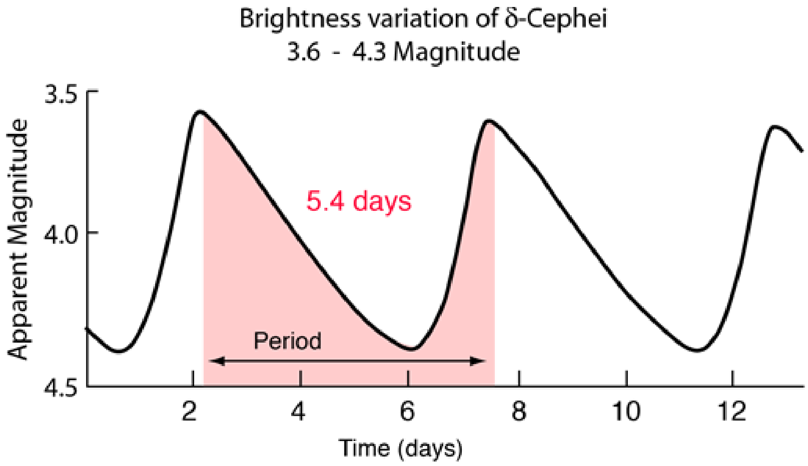
It achieved mean squared loss of 0.000315 in the training dataset, and a mean squared loss of 0.000567 in the testing dataset. The scatterplot roughly follows the goal red line of predicted redshift = actual redshift with a few outliers. However, there is a significant failure in galaxies with redshift < 0.05, and this is most likely because there is the relation between the input flux and redshift is too small in nearby galaxies.



[6] diagram generated from http://alexlenail.me/NN-SVG/index.html

Cepheid variable stars

Cepheid variable stars are a type of star that pulsates physically, changing in both diameter and brightness. In 1912, astronomer Henrietta Swan Leavitt discovered a relationship between the period and luminosity of Cepheid variables when investigating and cataloguing thousands of variable stars in the Small Magellanic Cloud and the Large Magallanic Cloud. In 1924, Edwin Hubble showed that the Cepheid variables in the Andromeda Galaxy were not in our own, disproving the then common view that the universe only consists of our Milky Way galaxy. Here is a typical light curve[7]:



Cepheid variables pulsates due to the kappa mechanism, where kappa represents the radiative opacity at a particular depth of the stellar atmosphere. In a normal star, an increase in compression of the atmosphere causes an increase in temperature and density, which decreases in the opacity of the atmosphere, allowing energy to escape more rapidly. The result is an equilibrium condition where temperature and pressure are maintained in a balance. However, in cases where the opacity increases with temperature (Cepheid variable stars), the atmosphere becomes unstable against pulsations. If a layer of a stellar atmosphere moves inward, it becomes denser and more opaque, causing heat flow to be blocked. In return, this heat increase causes a build-up of pressure that pushes the layer back out again. The result is a cyclic process in which the layer repeatedly moves inward and then is forced back out again, like steam leaking from a boiling kettle.

The paper *Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant*[8] is used in this essay for distances to other galaxies measured by Cepheid variable stars. It is calibrated as follows:

$$M_V = (-2.760 \pm 0.03)(\log_{10} P - 1) - (4.218 \pm 0.02)$$

$$M_I = (-2.962 \pm 0.02)(\log_{10} P - 1) - (4.904 \pm 0.01)$$

where P is the period in days, M_V is the absolute magnitude in visual V filter and M_I is the absolute magnitude in infrared I filter.

[7] from http://hyperphysics.phy-astr.gsu.edu/hbase/Astro/cepheid.html
[8] Wendy L. Freedman et al 2001

To account for interstellar extinction, the true distance modulus ($\mu = m - M$) is calculated:

$$\mu_0 = \mu_V - R(\mu_V - \mu_I)$$

where R is the reddening term, or the extinction factor.

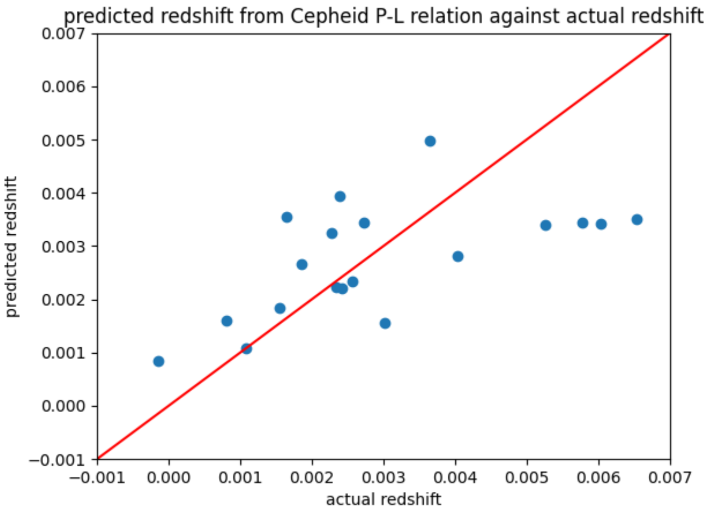
Since $m = -\sqrt[5]{100}\log_{10} F(d)$ and $M = -\sqrt[5]{100}\log_{10} F(10)$ where $F(d)$ is the flux observed from d megaparsecs away, the 2 equations can be combined into:

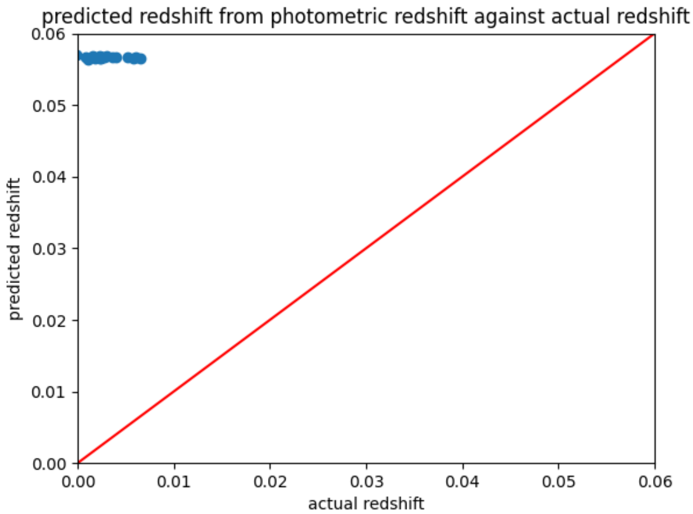
$$d = 10^{\frac{m-M}{5}+1} = 10^{\frac{\mu_0}{5}+1}$$

Finally, to obtain the corresponding redshift, the estimate $z \approx H_0 \times c \div d$ is used where H_0 is the Hubble constant and c is the speed of light.

$$z = H_0 c \times 10^{-\frac{\mu_0}{5}-1}$$

19 galaxies which had corresponding SDSS data were chosen from the 31 galaxies in the paper. The true distances were obtained from the SIMBAD astronomical database, where data of celestial objects are constantly updated.





The distance derived from Cepheid variables are accurate, while those produced by the neural network is completely off the mark. This is expected as the neural network performed poorly in small redshifts, and Cepheids can only be observed from close distances.

Type Ia supernovae

Supernovae are cataclysmic nuclear explosions occurring at the death of stars, they are very luminous objects able to outshine an entire galaxy, but later fade off slowly with time. A type Ia supernova is a subclass of supernovae which does not contain hydrogen, and presents a singly ionized silicon (Si ii) line at 615 nm near peak light. Type Ia supernovae are formed in binary star-white dwarf systems, where the gas from the star is transferred to the white dwarf due to gravity, the white dwarf accretes too much material and explodes as it reaches the Chandrashekar limit of 1.4 solar masses, the maximum stable mass of a white dwarf star.

Since all type Ia supernovae explode at almost the same conditions, it is supposed that the differences in peak luminosities of type Ia supernovae are correlated with how quickly their light curves decline after maximum light. The peak of the light curve of all type Ia supernova reaches a consistent luminosity with absolute magnitude $M_V = -19.3$, therefore they are said to be standard candles (a known standard luminosity).

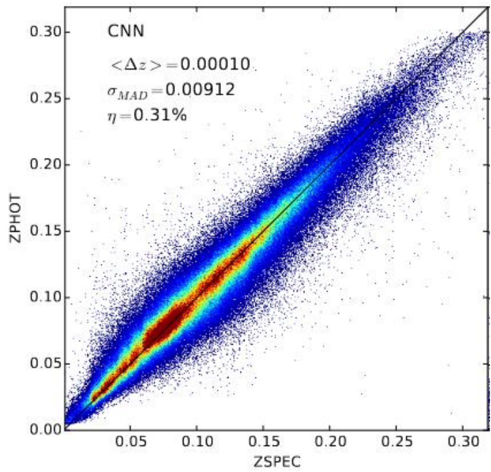
Since the absolute luminosity of type Ia supernovae is known, its distance can be calculated by measuring its apparent brightness.

[unfinished; couldn't extract host galaxy data from supernovae database]

Conclusion

In this essay I have trained my own neural network that produces photometric redshift. I then compared it to other methods of measuring distances to other galaxies by picking a galaxy set for each comparison and using that set to test the neural network and the other one.

It is found that for galaxies with small redshift, the period-luminosity relation of Cepheid variables perform considerably better than the photometric redshift. This tells us that photometric redshift should not be used for close galaxies. However, photometric redshift still have a few advantages compared to Cepheid variables: it can apply to all types of galaxies, not just galaxies containing a specific type of star; it can find the redshift of faraway galaxies while Cepheid variable stars in galaxies with large redshift are undetectable.



My neural network is far from perfect, and many optimizations can be made, such as considering galaxy morphology as inputs, or even using 5 64 by 64 images in the 5 filters as inputs, as shown on the right[9]. However, it demonstrates the basic concept of photometric redshift, and successfully capture the trend.

Moreover, more methods of measuring extragalatic distances can be compared, such as the Tully-Fisher relation in spiral galaxies and the Faber-Jackson Relation in elliptical galaxies. This essay serves as a basic exploration to measuring distances and photometric redshift, thus the old methods of Cepheid variables and type Ia supernovae were chosen.

[9] Johanna Pasquet et al, 2018

References and bibliography

Cappelari 2002, “Efficient multi-Gaussian expansion of galaxies”

Burke 2013, “The Assembly of Stellar Mass in Galaxy Cluster Cores over Cosmic Time”

Strauss, Michael A. et al 2002, “Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample”

Wendy L. Freedman et al 2001, “FINAL RESULTS FROM THE HUBBLE SPACE TELESCOPE KEY PROJECT TO MEASURE THE HUBBLE CONSTANT”

Johanna Pasquet et al, 2018, “Photometric redshifts from SDSS images using a Convolutional Neural Network”

John Yue Han Soo 2018, “ENHANCING PHOTOMETRIC REDSHIFTS FOR THE ERA OF PRECISION COSMOLOGY”