

Winter in Data Science

Data-Driven Astronomy

Harendra Pratap Singh Baghel
210040054

Content

INTRODUCTION AND LEARNINGS

1. What is data driven astronomy?
2. Introduction to Python
3. NumPy
4. Matplotlib
5. AstroPy
6. Web Scrapping
7. SciPy
8. image Reduction
9. Time Domain Astronomy
10. DDA- Module 1 to 6

What is data driven astronomy?

INTRODUCTION AND LEARNINGS

Data-driven astronomy (DDA) generates astronomical knowledge based on archived data sets, which may or may not be directly associated with the research in question.

In this project I learnt about different tools and that used in data science for reaseach and analysis how data is mearged into diffrent files for further usage.

I learned Python, Python libraries NumPy, SQL and many astronomical concepts like dwarf planet, super nova, black hole, planetary motion, galaxies, etc. and how data from various sources is processed and conclusions are made.

Introduction to Python

LEARNINGS

- Learned basics of python language required for programming.
- Features of python like:
 1. List
 - 2.Tuples
 - 3.Dictionary
 - 4.Loops and Conditions
 - 5.Functions
 6. Parsing a file for data

Introduction to NumPy

LEARNINGS

- Numpy arrays are similar to lists and tuples etc. in that they hold multiple values. However, all elements in a numpy array must be of the same data type.

- NumPy syntaxes with various functions:

1. `x=np.array([1,2,3,4])` # np.array converts any array-like object to a numpy array
2. `x = np.arange(0,100, 1)` # creates array of mentioned size
3. `y=np.copy(x)` # copy variable
4. `x = np.loadtxt('x.csv', delimiter=',')` # to load text file in the program

and various other functions of NumPy library

Matplotlib

LEARNINGS

- Matplotlib library helped to convert data into graphical format as graphs convey more than just raw tables.
 - It gave so many options to create different types of graphs like histograms, bar graphs, line graphs, scatter plots, etc.
 1. Syntaxes to plot different types of graphs: plt.plot([])- to plot point
 2. To mark the axes: plt.xlabel('some numbers')
 3. To make legend for the plot: plt.legend()
 4. To make separation by different colors of the plot and drawing formats:
- and various other features of Matplotlib.

Astropy

LEARNINGS

- Came to knew about most popular file format used in astronomy, Flexible Image Transport System (FITS), a didgital file format meant for storage of data.
- Came to know about telescope used to capture images in space called CCD camera. the array of numbers that is stored in a FITS file is basically a map of how many photons are incident on each pixel.

Web Scrapping

LEARNINGS

- Web scrapping is a technique for extracting data from websites and storing it in a file on your computer.
- Different files that a website contains are:
 - 1.HTML : Has the main contents of the page
 - 2.CSS : Used to "style" the webpage and make it look good
 - 3.JS : Javascript files make the webpage more interactive
 - 4.Images : Helps add images to the webpage
- How BeautifulSoup library is used for arranging webscrapping data.

SciPy

LEARNINGS

- There are almost always undetermined quantities present. G in Newton's law on Gravitation, the Planck's constant in Black Body Spectrum, Avogadro's Number, these were all just constants in theories.
- They all were determined experimentally, and the predictions made helped convert these theories into laws of physics.
- The `optimize.curve_fit` function of SciPy lets you do exactly that.

Image Reduction

LEARNINGS

- Understood the working of CCD Cameras used in astronomy and the working principle behind them. These cameras use capacitors and their property of bias voltage.
- Process of converting a raw image into a "useful" image is known as Image Reduction.
- Learnt ways of image reduction.

Time Domain Astronomy

LEARNINGS

- Branch of Astronomy which deals with objects that change with time.
Changes can be for various reasons.
- Discussed transients:
 1. Cepheid Variables, in particular, delta Cepheids
 2. Planetary Transits (an Exoplanet detected by Kepler K2)
 3. A type of transient that is all together not very well understood, but is extremely relevant in the current research in Transient Astronomy,
Gamma Ray Bursts

DDA- Coursera

- After going through the tutorials, there was a Coursera course on data-driven astronomy.
- MODULE 1- Thinking about data
 1. This module introduces the idea of computational thinking and how big data can make simple problems quite challenging to solve.
 2. the example of calculating the median and mean stack of a set of radio astronomy images illustrates some of the issues you encounter when working with large datasets.
 3. Learned about astronomical topics - Pulsars

DDA- Coursera

MODULE 2 : BIG DATA MAKES THINGS SLOW

- In this module, we explore the idea of scaling the code. Some algorithms scale well as your dataset increases, but others become impossibly slow.
- We look at some of the reasons for this and use the example of cross-matching astronomical catalogs to demonstrate what kind of improvements we can make.
- Astronomical Concept - Supermassive Black Hole

DDA- Coursera

MODULE 3 : QUERYING YOUR DATA

- Most large astronomy projects use databases to manage their data. In this module, we introduce SQL - the language most commonly used to query databases.
- We use SQL to query the NASA Exoplanet database and investigate the habitability of planets in other solar systems.
- Astronomical Concept - Exoplanets

DDA- Coursera

MODULE 4 : MANAGING YOUR DATA

- This module introduces the basic principles of setting up databases. We look at how to set up new tables and then how to combine Python and SQL to get the best out of both approaches.
- We use these tools to explore the life of stars in a stellar cluster.
- Astronomical Concept - Lifecycle of Stars

DDA- Coursera

MODULE 5 : LEARNING FROM DATA: REGRESSION

- This module introduces the idea of machine learning.
- We look at standard methodology for running machine learning experiments and then apply this to calculating redshifts of distant galaxies using decision trees for regression.
- Astronomical Concept - Cosmological Distances

DDA- Coursera

MODULE 6 : LEARNING FROM DATA: CLASSIFICATION

- In this module, we explore the limitations of decision tree classifiers.
- We then look at ensemble classifiers, using the random forest algorithm to classify images of galaxies into different types.
- Astronomical Concept - Types of galaxies and their classification

- Last, I want to conclude that this project was really interesting to do. Learned new concepts of space and how we are exploring the unexplored part of space with the help of science and technology.
- How raw data from different sources is collected and merged to get detailed results.
- I really want to thank the mentors for bringing this project to WiDS and selecting me for the project. The project helped me to dive into the ocean of data science and concepts of space and astronomy
- Really enjoyed doing the project :)