

Assignment # 1

Full marks-50, Deadline of submission: 17-10-2022

Hints on Solving the First Problem in ML (Data Collection) in Python:

1. Web Scrapping to collect COVID-19 data from
<https://www.worldometers.info/coronavirus/country/india/>

2. Web Scrapping to collect hotel data from the website

tripadvisor.in

1. Libraries you would need:
 - a. Numpy : to handle arrays
 - b. Pandas: library to perform data manipulation, you will be using to update and transform your data to store it in your result.
 - c. Urllib: library to perform url actions, access website etc.
 - d. BeautifulSoup: parsing library to extract data from html tags
2. Select your data collection website
3. Go through its pages, data you would collect
4. Analyze APIs you will be hitting, html tags that contains your data. MAKE A LIST OF IT.
5. Collect data in a tabular form (in csv or any text file). If you are storing images, collect it in a folder. For example: product details from flipkart.com store product images as file name: dove_shampoo_1 that is <product_name>_<product_category>_<img_number>
6. Try adding label to the data from collected information so that we have annotated data of our own.
7. Play around with it as much as you can and share information retrieved.

Note: Keep in mind data is to be collected which will help implement some algorithm to analyze and predict further information. For example, for COVID-19 data: design a predictor for forecasting the third wave characteristics, for hotel data: collect hotel name, price, discount, location, description and images (if possible)

Some other links and data repository can be found here:

- Goibibo.com/robots.txt
- Traveltriangle.com
- Or any other of your choice with justification
- MNIST data repository
- Kaggle Datasets
- UCI Machine Learning repository
- Amazon data set

35m

3. Following is the series of n natural numbers:

$$S=1+2+3+4+ \dots +n$$

- a) Calculate the series using formula, iterations and recursion methods and experimenting with different values of n , figure out the computational complexity(big O) in each case.
- b) When $n \rightarrow \infty$ can the above mentioned methods work? If not, can you find one solution ? 15m