MEENAKSHI -ASSIGNMENT1

In the second section, I learnt how to work with Numpy and how to add a particular library function to Python project interpreter. Numpy has a lot of inbuilt functions for creating and manipulating numpy arrays which include :-

1. array()
2. arrange()
3. ones()
4. shape()
5. zeros()

Scalar operations can be performed on the arrays easily using the operators \*, \*\*, -, +, 1/array.

Manipulation of arrays can be done easily using index values which can be accessed using “[]” and the slicing operator “:” for multiple indexes for 1D and 2D arrays . The indexes start from “0” for the first element and goes till “n-1”. As assigning values to a new array from an existing array creates only reference, the copy function is used to create a new unreferenced array. Another way of accessing the rows is by using the method “array[[n1,n2]]” which prints the ‘n1+1’th and ‘n2+1’th rows of the array. Loops such as “for“ can also be used for indexing the elements.

Some of the functions used for manipulating arrays include :-

1. “numpy.exp(A)”
2. “numpy.add(A,B)”
3. “numpy.sqrt(A)”
4. “numpy.maximum(A,B)”

Arrays can be saved in files and used whenever required.

1. “numpy.save(‘file\_name’,array\_name)” and “numpy.load(‘file\_name.npy’)” for single array.
2. “numpy.savez(‘Archive\_name.npz’, name1 = a1, name2 = a2,… )” can be used to save more than one array in an archive file.
3. “numpy.savetxt(‘textfile\_name’,array,delimiter = ‘ ’)” and “numpy.loadtxt(‘textfile\_name’,delimiter = ‘ ’)” can be used to save and load arrays in a text file where delimiter is the character by which the elements are separated in the text file.

Two 2d arrays can be used to plot graphs by using the mathplotlib library and using the functions imshow(fn) where fn is the function involving the arrays, titlebar(“String”), colorbar(), savefig(‘name.png’) .

Conditions can be used to assign and manipulate array elements using the function “where(condition,array1,array2)”. The condition is tested and if true, element of array1 is selected else element of array2 is selected.

In the third section, introduction to the library “Pandas” was given. Pandas has different data structures such as series and dataframes . Series is an array like structure where the index of each element can be changed as required. Series can be created by using the code “Series(numpy\_array)” which converts the numpy array into a series. Another argument called the index array can be used to name the indexes which will have a null value for an extra index.

DataFrame is a 2d array like structure with a specific row name and column name. Thus Dataframes can be used to create a table of contents. Some of the functions on dataframe df are :-

1. df.columns
2. df.head
3. df.tail
4. df.ix[n]
5. df[column\_name]
6. df[row\_name]
7. del df

Reindexing is the process of setting new indexes to arrays, series etc. Series can be reindexed by using the function “reindex([new\_indexes])” ,the method fill\_value is used to assign a value to all the new indexes, the method ffill is used to fill the indexes if they are not continuous.

Dataframes can be reindexed by using reindex([row\_indexes]), reindex(columns = [column\_indexes]), ix([row\_index], [column\_index]) .

The drop function is used to delete a particular row or column. drop(index) removes the specified row of the series or a dataframe, drop(index, axis = 1 ) deletes the column of the dataframe.

The null data is specified as NaN. We have to handle the null data to avoid ambiguity and errors in the code. “dropna” deletes the rows with null data and “dropna(how=’all’)” deletes the row if all the elements are null. “dropna(axis=1)” deletes the column and “dropna(thresh=3)”.

The “[]” and “:” can be used to access the values of series and dataframes. If a dataframe consists of null values then during addition with another dataframe the sum of the elements are given null values. Hence “df1.add(df2,fill\_value=n)” can be used where df1 and df2 are dataframes and the value n is given to the null values.

Each value of a series has an index and rank. The series can be sorted based on the index and values. The rank is given based on the ascending order of the values.

The elements of the Dataframes can be plotted in the form of a graph using the Mathplotlib library.