DATA SCIENCE AND MACHINE LEARNING

MANU SINGH

([manusingh221101@gmail.com](mailto:manusingh221101@gmail.com))

Task 1

( Pandas and NumPy)

**Pandas**

**Definition:**  It is most widely used for machine learning and data science in which this library helps to read data as well as helps to analyze the different formats of the data.

Pandas allows various data manipulations such as merging, reshaping selecting as well as cleaning data.

The name pandas has a reference to “python data analysis” and was created by Wes McKinney in 2008.

**Getting started with pandas**

Import pandas:

when we start working on any project we need to read our data so importing pandas library is very important.

import pandas

student= {‘name’:[“ manu”,”shagun”,”tanu”], ‘passing’ : [2,1,3]}

myvar = pandas.DataFrame(student)

print(myvar)

**What pandas can do ?**

Pandas gives us answers about the data like : average value ,max value, min value and the correlation between two rows or columns. Pandas also abe to delete rows that are not relevant, or contains wrong values, like empty or NULL values called as cleaning the data.

**Pandas series**

It is a one dimensional array holds data of any type.

**Labels:** The index number are used to label the values if nothing are specified . It becomes very easy to access the values from any data by using its index number.

import pandas as pd

x=[1,2,3]

value =pd. Series (x, index = [“a”,”b”,”c”])

print(value)

#to access value of b

Print(value[“y”])

**DataFrame:** multidimensional tables are called DataFrames. Series is a column but DataFrame is the whole table.

import pandas as pd

data= {“mango”:[1,2,3],”apple”:[10,20,30]}

#load data into DataFrame object:

Df =pd.DataFrame(data)

Print(df)

**Naming and locating Index :** you can also name your index.

Add a list of names to give each row a name.

Import pandas as pd

data= {“mango”:[1,2,3],”apple”:[10,20,30]}

df = pd.Dataframe (data,index= [“ price1”,”price2”,”price3”])

print(df)

#refer to the named index:

print(df.loc[“price3”])

**Load files into a DataFrame**

Import pandas as pd

df = pd.read\_csv(‘data.csv’)

print(df)

**Pandas – Cleaning Data**

**Data cleaning** means fixing bad data (bad data like: empty cells,data in wrong format,duplicates) in your data set.

**Remove Rows**

Import pandas as pd

df= pd.read\_csv(‘data.csv’)

new\_df = df.dropna()

print(new\_df.to\_string())

# dropna() method returns a new DataFrame and will not change the original.

**Remove all rows with null values**

df=pd.read\_csv(‘data.csv’)

df.dropna(inplace = True)

print(df.to\_string())

**replace empty values**

df.fillna(10, inplace = True)

**replace only for a specified columns**

df[“apple”].fiilna(10,inplace = True)

**replace using mean, median or mode**

1.x =df[“apple”].,mean()

2.x =df[“apple”].median()

3.x = df[“apple”].mode()[0]

**NUMPY**

Its is used to working with arrays in python library and we can use it freely. NumPy was created in 2005 by Travis Oliphant.

NumPy arrays are stored at one continuous place in memory unlike lists, so process can access and manipulate them very efficiently. The most of the part of the NumPy are written in C or C++.

CREATE A NumPy ndarray OBJECT

The array object in NumPy is called ndarray.

import numpy as np

arr = np.array([1,2,3,45])

print(arr)

print(type(arr))

# 0-D array

import numpy as np

arr= np.array(42)

print(arr)

# 1-D array

arr1= np.array([1,2,3,4])

print(arr1)

# 2-D array

arr2 = np.array([[1,2,3,4] , [5,6,7,8]])

print(arr2)

#3 D array

arr3 = np.array([[[1,2,3,4] , [5,6,7,8], [1,2,3,4]]])

print(arr3)

# to check the number of dimensions

print(arr0.ndim)

print(arr1.ndim)

print(arr2.ndim)

print(arr3.ndim)

NumPy DATA TYPES

* Strings
* Integer
* Float
* Boolean
* Complex
* Object
* Unicode string
* Timedelta

#checking the datatype of an array

Import numpy as np

arr= np.array([1,2,3,4,])

print(arr.dtype)

**NumPy array copy vs view**

The main difference between a copy and a view of an array is that the copy is a new array, and the view is just a view of the original array.

**#copy array**

import numpy as np

arr = np.array([1,2,3,4,5])

x= arr.copy()

arr[0]=42

print(arr)

print(x)

**# view array**

import numpy as np

arr = np.array([1,2,3,4,5])

x= arr.view()

arr[0]=42

print(arr)

print(x)

**NumPy SEARCHING ARRAYS**

Import numpy as np

Arr= np.array([1,2,3,4,5])

x =np.where(arr ==4)

print(x)

#**find the indexes where the values are even**

Import numpy as np

Arr= np.array([1,2,3,4,5])

x =np.where(arr %2==0)

print(x)

#**find the indexes where the values are odd**

Import numpy as np

Arr= np.array([1,2,3,4,5])

x =np.where(arr %2==1)

print(x)

**NumPy filter array**

Import numpy as np

arr = np.array([1,2,3,4,5])

x= [True, True, False, True, False]

newarr = arr[x]

print(newarr)