Data Science with Python

Why Python

1. Easy to learn
2. Not just statistics language but data acquisition, cleaning, databases, high performance computing and modeling
3. Strong data science libraries eg Scipy ecosystem

Python pandas toolkit

Advanced querying and manipulation. Basic statistical analysis with numpy and scipy

Good data science brings skepticism, experimentation, simulation and replication.

Python is high level language, interpreted directly compiled into machine code. Good for where exploration is better, dynamically typed language. Set and modify the variables. Python interpreter is stateful.

Static typing in python – strings

Tuples – immutable, mixed types, unpacking in python

Lists – mutable, mixed types, + sign concatenates lists, \* operation repeats elements in the lists, all strings are lists of characters, Slicing in python

Dictionaries – iterate with items function

%precision 2

With open(‘file.csv’) as csvfile:

List(csv.DictReader(csvfile))

Data science with python continued

Jan 1st 1970 epic time

Import datetime as dt

Import time as tm

Tm.time()

Dtnow = dt.datetime.fromtimestamp(tm.time())

dtnow.year, dtnow.month , how , minute, second

dt.timedelta(100)

Object oriented python

class Person:

dept = “ssdsdf”

def set\_name(self,new\_name):

## do something

Implications of oops in python

Dont have private and protected members

No need for explicit constructers in python

Map function is basis for functional programming in python

map(function, iterable), it helps for lazy evaluation and so does not allow till we try to look at the value

Advanced python lambda and list comprehensions

Lambda are anonymous functions and they don’t have name

my\_function = lambda a,b,c : a+b

Very useful for data cleaning tasks

for person in people:

    print(split\_title\_and\_name(person) == (lambda x: x.split()[0] + ' ' + x.split()[-1])(person))

#option 2

list(map(split\_title\_and\_name, people)) == list(map(lambda person: person.split()[0] + ' ' + person.split()[-1], people))

List comprehensions

numpy for matrices

Bumpy arange, resize, np.linespace,eye,diag,ones,repeat,vstack,hstack

+,-,\*\*,\*,dot function,dtype,astype,sum,max,min,mean,std,argmax(),argmin(),

Indexing and slicing

R2 = r[:3,:3]

If we make any changes to the values in r2 then automatically it will be reflected in the corresponding section of r as well

len(), range(), enumerate functions to iterate over the array

Zip to iterate over both arrays in together