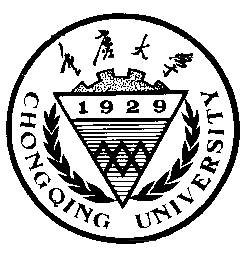
数据科学导论实验报告

实验一 python数据基础实践



学 生：李嘉迪\_20191737

年 级：2019级

专 业：数据科学与大数据技术

重庆大学大数据与软件学院

2020年 04月 19日

## 实验目的：

为后续上机实验做准备，掌握 python 基础操作：包括熟悉 python 程序运行环 境及基本语法，理解 python 基本数据类型，掌握 python 的算术运算规则及表达 式的书写方法；掌握常用的 python 基本数据函数分析库：pandas、numpy。

## 实验要求

工具：anconda/spyder/pycharm 语言：python

## 开发环境：

Windows 10

## 实验内容：

**Python 基础操作**

（1）读入 2 个正整数 A 和 B，1<=A<=9, 1<=B<=10,产生数字 AA…A,一共 B 个 A（例如输入 2 和 3，输出为 222）

（2）将字符串 str = "hElLo pYtHOn "做以下变换： 1）把所有字符中的小写字母转换成大写字母； 2）把所有字符中的大写字母转换成小写字母； 3）把第一个字母转化为大写字母，其余小写； 4）把每个单词的第一个字母转化为大写，其余小写

（3） 输出 10 个不重复的英文字母 (字符串转列表实现)：随机输入一个字符 串，把最左边的 10 个不重复的英文字母（不区分大小写）挑选出来。 如 没有 10 个英文字母，显示信息“not found”

（4）给定一个字符串，判断它是否是回文字符串（即类似于 level, 12321 这样 的对称字符串），如果是输出回文及 YES，不是则输出原始字符串及 N O。 （5）小明身高 1.75，体重 80.5kg。请根据 BMI 公式（体重除以身高的平方） 帮小明计算他的 BMI 指数，并根据 BMI 指数： 低于 18.5： 过轻 18.5-25： 正常 25-28： 过重 28-32： 肥胖 高于 32： 严重肥胖 用 if-elif 判断并打印结果：

（6）输入一个数字： 1）若该数字各个位数之和为奇数，则将该数各个位数倒叙打印（如 12 2，打印为 221）。 2）若该数字个位数之和为偶数，则直接打印该数字。

（7）编写程序，输入两个集合 setA 和 setB,分别输出它们的交集，并集和差 集

（8）编写程序：输入一个自然数，输出它的二进制，八进制，十六进制表示形式

（9）编写程序：输入一个包含若干整数的列表，输出一个新列表，要求新列 表中只包含原列表的偶数(如请输入一个列表：[1,2,3,4,5,6]输出[2, 4, 6])

（10）编写程序，输入两个分别包含若干整数的列表 lstA 和 lstB,输出一个字 典，要求使用列表 lstA 中的元素作为键，列表 lstB 中的元素作为值，并且 最终字典中的元素数量取决于 lstA 和 lstB 中元素最少的列表的数量。（如 请输入 listA 列表：[1,2,4,6]；请输入 listB 列表：[3,5,6,8]输出{1: 3, 2: 5, 4: 6, 6: 8}）

（11）假设已有列表 A=[(“dog”,“type”),(“black”, “color”),(“cat”,“type”),(“blue”, “color”),(“green”,“color”),(“pig”,“type”)]，其中每个元素都是一个元组。元 组中的第一个元素表示值，第二个元素表示标签。试编写程序，将所有的 颜色值从列表 A 中提取出来，存入列表 A\_colors,并将该列表输出。

（12）输入数字计算阶乘：整数的阶乘（英语：factorial）是所有小于及等于 该数的正整数的积，0 的阶乘为 1。即：n!=1×2×3×...×n。

（13）编程实现：输入的数字是否为阿姆斯特朗数。（如果一个 n 位正整数等 于其各位数字的 n 次方之和,则称该数为阿姆斯特朗数。 例如 1^3 + 5^3 + 3^3 = 153。1000 以内的阿姆斯特朗数： 1, 2, 3, 4, 5, 6, 7, 8, 9, 153, 370, 371, 407。）

（14）对以下元素 88,5,23,31,45,4,6,1,16,12 进行冒泡排序

（15）编程实现五人分鱼：A、B、C、D、E 五人在某天夜里合伙去捕鱼，到 第二天凌晨时都疲惫不堪，于是各自找地方睡觉。日上三杆，A 第一个醒 来，他将鱼分为五份，把多余的一条鱼扔掉，拿走自己的一份。B 第二个 醒来，也将鱼分为五份，把多余的一条鱼扔掉拿走自己的一份 。C、D、 E 依次醒来，也按同样的方法拿鱼。问他们至少捕了多少条鱼?

Python 基本函数分析库

（一） 熟悉常用的 pandas 操作

(1) 用 pandas 中 Series 创建一个序列 s 为[1,2,3,4],标签为[a,b,c,d]

(2) 用 DataFrame 创建一个二维表 d a b c 1 2 3 4 5 6 ;也可用已有的序列来创建表格

(3) 用 pandas 读取文件 ocean\_temp.csv,命名为 df，注意:文件的存储路径不 能带有中文，否则读取可能出错; 读取文本格式的数据，一般用 encoding 指定编码 utf-8。

(4) 预览 df 前 5 行数据

(5) 统计 df 数据基本统计量(count, mean, std, min, quantiles, max)

(6) 在 df 的基础上创建一个从 10 到 50 行（包含 10 行和 50 行）的片段 df\_slice，并输出。

（二）熟悉常用的 Numpy 操作

(1) 从列表 df 中创建一个名为 tempArr 的 NumPy 数组（ndarray）,然后输出 tempArr 的大小和类型。

(2) 创建另一个与 tempArr 大小相同的数组 adjar，但这次所有元素都设置为 20

(3) 将上述两个数组添加到一个新的数组中，新数组命名为 newTemp

(4) 在 newTemp 中的温度是华氏度。定义一个函数，将温度从华氏度转换为摄 氏度，然后使用矢量化函数将 newTemp 中的所有温度数据转换为摄氏度， 并存储在新数组 cTemp 中。 华氏度转摄氏度公式如下： ( 32) 9 5 C= F− (5) 显示上述所有数组的前 50 个元素

(6) 显示上述 cTemp 数据的 20 到 50 行

(7) 创建基本的 10x10 数组：1）所有元素都是 0；2）I 矩阵（即对角线 0 上 的值 1）；3）所有元素都是在[0,100]范围内随机生成的整数。输出所有 数组。

## 实验结果

Python基础操作

（1）

a = input("plz enter number (1<=A<=9):")

b = input("plz enter number (1<=B<=10):")

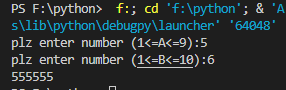
while(int(a)<1 or int(a)>10 or int(b)<0 or int(b)>11 ):

    a = input("plz enter number (1<=A<=9):")

    b = input("plz enter number (1<=B<=10):")

for i in range(0,int(b)):

    print(a,end="")



（2）

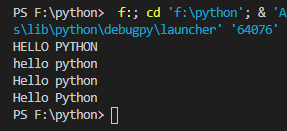
str = "hElLo pYtHOn"

print(str.upper())

print(str.lower())

print(str.capitalize())

print(str.title())



（3）

sentence = input()

str = list(sentence)

temp = [0,0,0,0,0,0,0,0,0,0]

for i in range(0,10):

    temp[i] = str[i]

new\_list=[]

for i in str:

    if i not in new\_list:

        new\_list.append(i)

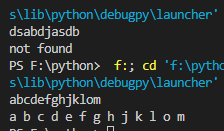
if len(str)==len(new\_list):

    for i in str:

        print(i,end=" ")

else:

    print("not found")



（4）

s = input('')

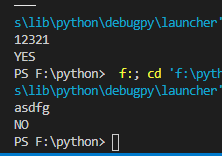
a = reversed(list(s))

if list(a) == list(s):

    print('YES')

else:

    print('NO')



（5）

heigh = 1.75

weight = 80.5

BMI = 80.5/(1.75\*1.7)

if(BMI<18.5):

    print("过轻")

elif(18.5<=BMI<25):

    print("正常")

elif(25<=BMI<28):

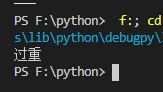
    print("过重")

elif(28<=BMI<32):

    print("肥胖")

else:

    print("严重肥胖")



（6）

i = input("plz enter a number: ")

str = list(i)

temp = []

sum = 0

for j in str:

    sum = sum + int(j)

if (sum%2!=0):

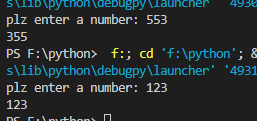
    temp = reversed(list(str))

    for j in temp:

        print(j,end="")

else:

    print(i,end="")



（7）

A = input("enter setA: ")

B = input("enter setB: ")

setA = A.split()

setB = B.split()

print("setA & setB")

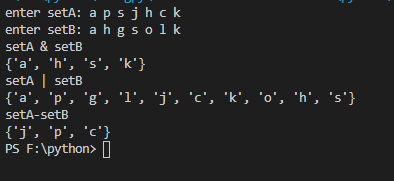
print(set(setA)&set(setB))

print("setA | setB")

print(set(setA)|set(setB))

print("setA-setB")

print(set(setA)-set(setB))



（8）

num = int(input(""))

num\_2 = bin(num)

num\_8 = oct(num)

num\_16 = hex(num)

print(num,num\_2,num\_8,num\_16)



（9）

num = input("")

num\_list = num.split(" ")

new\_list = []

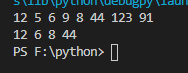
for i in num\_list:

    if(int(i)%2==0):

        new\_list.append(i)

for i in new\_list:

    print(i,end=" ")



（10）

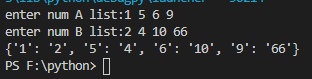
A = input("enter num A list:")

B = input("enter num B list:")

A\_list = A.split(" ")

B\_list = B.split(" ")

print(dict(zip(A\_list,B\_list)))



（11）

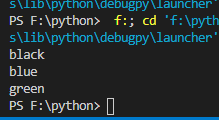
A=[("dog","type"),("black","color"),("cat","type"),("blue","color"),("green","color"),("pig","type")]

B = dict(A)

for key,val in B.items():

    if val == 'color':

        print(key)



（12）

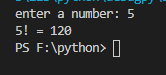
num = input("enter a number: ")

sum = 1

for i in range(1,int(num)+1):

    sum = sum\*i

print("%d! = %d"%(int(num),sum))



（13）

num = int(input("enter a num:"))

order = len(str(num))

sum = 0

for i in str(num):

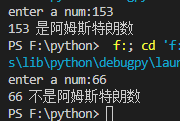
    sum = sum + int(i)\*\*order

if num == sum:

    print(num, "是阿姆斯特朗数")

else:

    print(num, "不是阿姆斯特朗数")



（14）

def sort(arr):

    n = len(arr)

    for i in range(n):

        for j in range(0, n-i-1):

            if arr[j] > arr[j+1] :

                temp = arr[j]

                arr[j] = arr[j+1]

                arr[j+1] = temp

arr = [88,5,23,31,45,4,6,1,16,12]

print ("原数组:")

for i in arr:

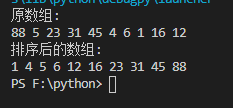
    print (i,end=" ")

sort(arr)

print ("\n排序后的数组:")

for i in arr:

    print (i,end=" ")



（15）

fish = 1

while True:

    total = fish

    flag = True

    for i in range(5):

        if (total-1) % 5 == 0:

            total = (total - 1) /5 \* 4

        else:

            flag = False

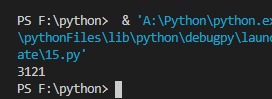
            break

    if flag:

        print(fish)

        break

    fish = fish + 1



Python基本函数分析库

（一）Pandas操作

(1)

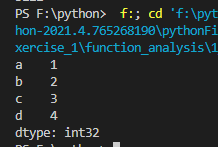
import numpy as np, pandas as pd

arr\_value = np.arange(1,5)

arr\_index = ['a','b','c','d']

s1 = pd.Series(arr\_value,index = arr\_index)

print(s1)



(2)

import pandas as pd

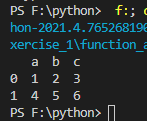
data = {'a':[1,4],

        'b':[2,5],

        'c':[3,6]}

list = pd.DataFrame(data)

print(list)



(3)

import pandas as pd

df = pd.read\_csv(r'F:\python\Data\_Science\_exercise\exercise\_1\function\_analysis\ocean\_temp.csv',encoding='UTF-8')

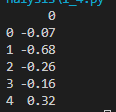
(4)

import pandas as pd

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

print(df.head(5))



(5)

import pandas as pd

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

print('count:%d'%df.count())

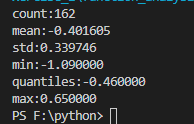
print('mean:%f'%df.mean())

print('std:%f'%df.std())

print('min:%f'%df.min())

print('quantiles:%f'%df.quantile())

print('max:%f'%df.max())



(6)

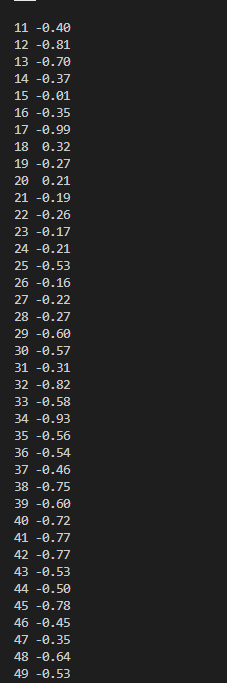
import pandas as pd

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

df\_slice = df.loc[10:50]

print(df\_slice)



Numpy操作

(1)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

tempArr = np.array(df)

print("大小：",tempArr.size)

print("类型：",type(tempArr))



(2)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

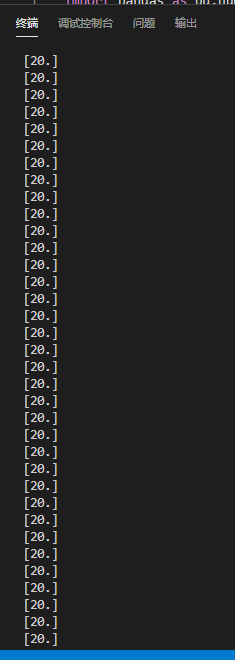
adjar = np.array(df)

print(adjar)

for i in range(len(adjar)):

    adjar[i] = 20.0

print(adjar)



(3)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

adjar = np.array(df)

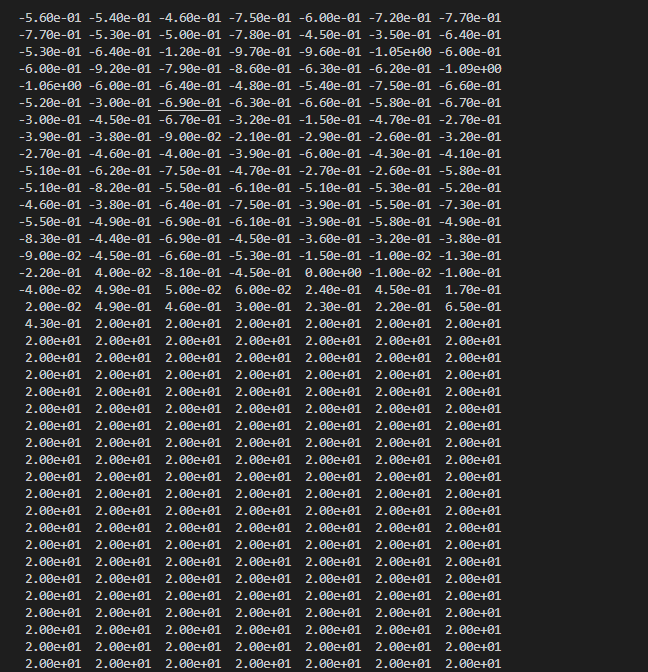
tempArr = np.array(df)

for i in range(len(adjar)):

    adjar[i] = 20.0

newTemp = np.append(tempArr,adjar)

print(newTemp)



(4)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

adjar = np.array(df)

tempArr = np.array(df)

for i in range(len(adjar)):

    adjar[i] = 20.0

newTemp = np.append(tempArr,adjar)

cTemp = []

def change(f):

    c = 5.0/9.0\*(f-32.0)

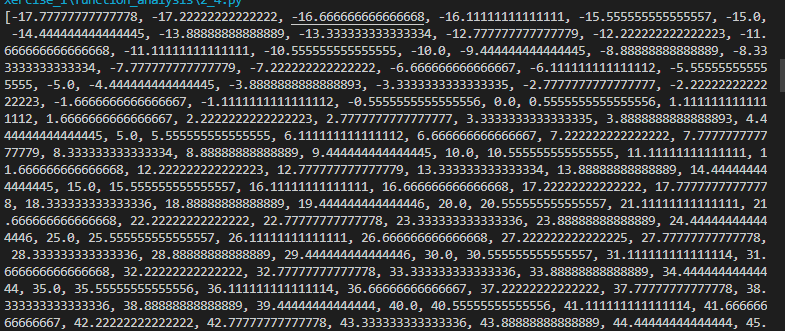
    return c

for i in range(len(newTemp)):

    temp = change(i)

    cTemp.append(temp)

print(cTemp)



(5)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

adjar = np.array(df)

tempArr = np.array(df)

for i in range(len(adjar)):

    adjar[i] = 20.0

newTemp = np.append(tempArr,adjar)

cTemp = []

def change(f):

    c = 5.0/9.0\*(f-32.0)

    return c

for i in range(len(newTemp)):

    temp = change(i)

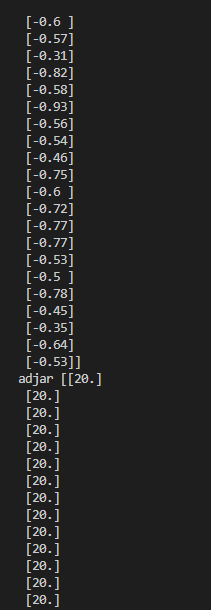
    cTemp.append(temp)

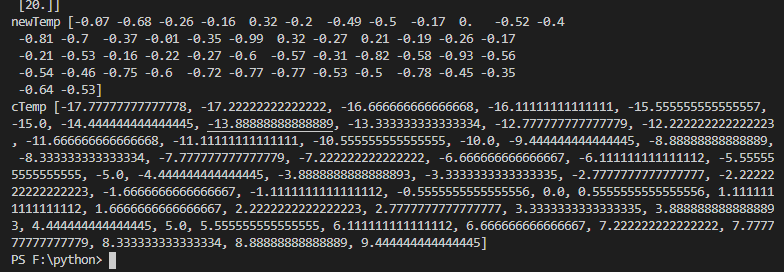
print("tempArr",tempArr[:50])

print("adjar",adjar[:50])

print("newTemp",newTemp[:50])

print("cTemp",cTemp[:50])





(6)

import pandas as pd,numpy as np

file = 'F:/python/Data\_Science\_exercise/exercise\_1/function\_analysis/ocean\_temp.csv'

df = pd.read\_csv(file,encoding='UTF-8',header=None)

adjar = np.array(df)

tempArr = np.array(df)

for i in range(len(adjar)):

    adjar[i] = 20.0

newTemp = np.append(tempArr,adjar)

cTemp = []

def change(f):

    c = 5.0/9.0\*(f-32.0)

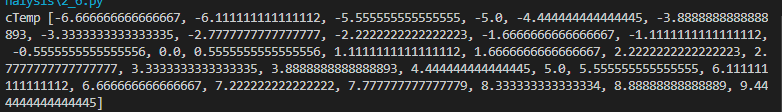
    return c

for i in range(len(newTemp)):

    temp = change(i)

    cTemp.append(temp)

print("cTemp",cTemp[20:50])



(7)

import numpy as np

arr1 = np.zeros([10,10],dtype=int)

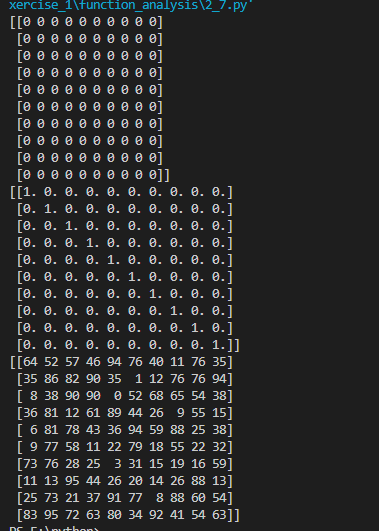
print(arr1)

arr2 = np.eye(10,k=0)

print(arr2)

arr3 = np.random.randint(0,100,size = (10,10))

print(arr3)



## 问题和解决