实验报告

东北林业大学

信息与计算机科学技术实验中心

|  |
| --- |
| 一、实验目的  （1）了解Numpy库的基本功能，包括数组的创建、切片、修改元素、排序等方法；  （2）掌握利用Numpy库中的通用函数对数组进行操作与运算的方法；  （3）掌握利用Numpy库对数据进行统计与分析的方法。。 |
| 二、实验环境  （1）硬件：PC机；  （2）软件：Anaconda Jupyter Notebook，Spyder，Numpy |
| 三、实验内容及结果  （一）数组的创建与基本操作  1、创建一个长度为10的一维全为0的ndarray数组，然后让第5个元素等于1。    2、创建一个5行3列的二维数组ar，元素大小为从[10-15)区间中随机数组成的整数。    3、查看2题中创建的数组ar的大小、元素个数、维度个数和数组元素的数据类型及数组元素的字节数。    4、将2题中创建的数组修改为3行5列后输出最后一行的元素。    5、将数组ar展平为一维数组，存储至ar2中，然后利用切片提取数组中从下标为1到倒数第二个元素之间的所有元素进行输出；    或者：    6、将ar2中下标为1的元素修改为100，然后输出ar和ar2数组中各元素的值，分析Numpy中数组切片与Python中列表切片的结果有什么区别。    7、对利用切片对ar2中的元素进行反序处理后输出。  8、创建一个二维数组，使用索引方式获取第二行第一列和第三行第二列的元素。    9、利用创建一个由从0开始到9范围间隔为1的整数组成的数组a，和一个从10到30范围内偶数组成的数组b，然后将a和b转换为3\*3的数组后做水平合并和垂直合并，观察合并后生成的数组元素，比较hstack和vstack函数的功能有什么不同。    10、创建一个存储6个同学姓名的数组names，和一个存放6名同学考试成绩的数组grades，保证每个同学的姓名和成绩的序号一致，接收用户输入的自己姓名，然后查询该同学的考试成绩。  其差别可以参看下图：    11、统计第10题中grades数组中的最大值、最小值、平均值、中位数信息，查看学生成绩的统计情况。  12、根据第10题中的grades数组中的百分制成绩生成二级制成绩单，两级分别为通过和不通过。  （二）利用numpy相关函数对鸢尾花数据集进行数据分析  1、导入isrs.npy文件中的数据  2、输出数据集中的数据，查看数据的维度、元素个数；  3、查看鸢尾花的类型有哪些，注意去除重复值；  4、获取x中的前4列数据存入data数组，获取最后一列存入flowers数组  5、查看鸢尾花数据集中各属性数据的最大值、最小值、平均值等信息；  6、查看Setosa鸢尾花、Versicolor鸢尾花和Virginica鸢尾花各类样本的数量  7、查看Setosa鸢尾花、Versicolour鸢尾花和Virginica鸢尾花花萼的最大值、最小值、均值等数据，分析不同类型鸢尾花数据的特征，选择区分度较大的属性作为后续工作的依据  （二）利用numpy相关函数对我国六普和七普人口数据进行数据分析    1、导入人口数据，数据包中的popu.npy文件是各省六普和七普人口数据,显示人口数据，其中第一列是行政区数据，第二列为2010年人口数据，第三列为2020年人口数据    2、统计我国2010年和2020年的总人口  3、将各省数据按照2010年数据降序排列。  4、将各省数据按照2020年数据降序排列  5、计算各省从2010到2020年的人口变化量，单位为万。  6、将gt中的人口变化数据与ar数组进行横向合并。   1. 数组行或列的插入函数   7、按照同样的方法计算各省份人口变化占2010年总人口的比例，并将该数组作为data的第4列插入到数组中。  8、查看2010-2020年期间人口增加的省份和人口减少的省份。  从中可以看出，人口净流入的城市主要为东北、西北和西部城市，中部城市仅有山西省一个省份。  9、将省份按照人口增长的幅度进行降序排列，分析2010-2020年期间各省份人口增长与区域经济、区位等因素的关系。  10、(选做）将结果用地图显示 |

|  |
| --- |
| 四、实验过程分析与讨论  1． 实验过程分析与讨论：  （1）在本实验中，主要了解Numpy库的基本功能，包括数组的创建、切片、修改元素、排序等方法；在第一个实验中，np.zeros()可以创建全是0的数组，与此同时，ar[n]可以将第n-1赋值。  （2）shape函数可以规定二维数组，几行几列，但是要提前给数组进行赋值，这个时候可以利用随机函数。  （3）ar.shape: 这个属性返回一个元组，表示数组的形状。ar.size: 这个属性返回数组中元素的总数。例如，如果一个数组是二维的，那么它的元素总数就是行数乘以列数。ar.ndim: 这个属性返回数组的维度数。ar.dtype: 这个属性返回数组中元素的数据类型。例如，如果一个数组中的元素都是整数，那么它的数据类型就是int。ar.itemsize: 这个属性返回数组中每个元素的字节大小。例如，如果一个数组中的元素都是32位整数，那么它的itemsize就是4。  （4）ar.ravel()用于将多维数组转换为一维数组。它返回一个新的一维数组，其中包含原始数组中的所有元素，按照原始数组的元素顺序排列。如果 ar 是一个空数组，即没有任何元素的数组，那么调用 ravel() 方法会返回一个空的一维数组。  （5）ar.flatten()用于将多维数组转换为一维数组。与 ravel() 方法不同的是，flatten() 方法总是返回一个新的一维数组，而不是返回原始数组的视图。这意味着 flatten() 方法会创建一个新的数组对象，其中包含原始数组中的所有元素，按照原始数组的元素顺序排列。  （6）ar[-1::-1]：这是一个切片操作，用于获取数组 ar 的逆序（倒序）子集。在这个例子中，-1 表示最后一个元素的索引，::-1 表示逆序切片的步长为 -1。  （7）names=np.array(['mary','ton','ketty','nike','lily','rose'])：定义了一个名为 names 的一维数组，其中包含了六个学生的姓名。使用布尔数组作为索引，从 grades 数组中获取与输入姓名相等的成绩。这里使用了 NumPy 的布尔索引功能，它允许我们使用布尔数组来选择数组中的元素。  （8）在第二组实验中，使用load打开npy文件，与此同时np.unique 是 NumPy 库中的一个函数，用于获取数组中的唯一值。在这个例子中，它被用于获取二维数组 x 最后一列的唯一值。  （9）在使用Map库函数时，要注意Map函数更新，在中国地图中，名称是省份全称，而不是缩写，所以在npy文件行1为省份缩写，导致地图由于没有读到数据，所以并不能通过不同数据得到不一样的颜色，全是灰色，所以应该把npy第一行的数据变成全称，可以采用先删除第一列，然后再插入全程的第一列。 |
| 五、指导教师意见    指导教师签字：  年 月 日 |