## 1.3 Python基本语法

Python是一种计算机编程语言。计算机编程语言和我们日常使用的自然语言有所不同，最大的区别就是，自然语言在不同的语境下有不同的理解，而计算机要根据编程语言执行任务，就必须保证编程语言写出的程序决不能有歧义，所以，任何一种编程语言都有自己的一套语法，编译器或者解释器就是负责把符合语法的程序代码转换成CPU能够执行的机器码，然后执行。Python也不例外。

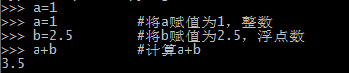
任何计算机系统，其工作的实质都在于，将输入后的数据处理后再输出，即：

输入→处理→输出

### 1.3.1数据类型——数字和字符串

计算机，顾名思义就是可以做数学计算的机器，因此，计算机程序理所当然地可以处理各种数值。但是，计算机能处理的远不止数值，还可以处理文本、图形、音频、视频、网页等各种各样的数据，不同的数据，需要定义不同的数据类型。数据类型的内容有很多，如字典、文件、列表和元组。

作为动态类型的语言，Python中使用数字无需先声明它的类型，如下所示。



在上面的例子中，我们分别将整数1和浮点数2.5赋值给了变量a和b，“=”就是赋值的意思，“+”为运算符，在Python中，除了加减乘除取余等运算符外，还有逻辑运算符如位移运算符等等。接下来直接输入公式即可完成计算。每句后面以“#”开头的文本时注释，它对于脚本的运行没有任何影响，目前可以完全被我们忽略掉——当然，在复杂脚本里面使用注释说明自己编程时的思路算法也是一个很好的习惯。

Python中的字符串是用来表示和储存文本的，通常由单引号、双引号或者三引号包围，可以包含数字、字母以及一些控制字符，如换行符（/n）、制表符（/t）、和回车（/r）等等。字符串的形式一般如下所示。

IMG_256

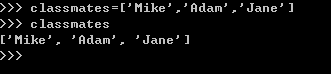
接下来用实例演示转义字符在字符串中的使用。



这个例子里面，print语句的作用是输出字符串，只有在使用print输出字符串时才能够解释字符串中的转义字符。需要注意的是，在Python3中，print的正确语法为print（），否则就会显示错误。

### 1.3.2 数据类型——列表list

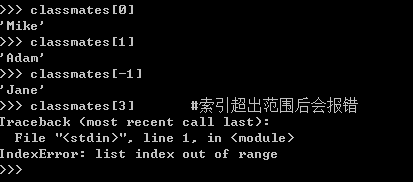
Python内置的一种数据类型是列表：list。List是一种有序的集合，可以随时添加和删除其中的元素。比如，列出班里所有同学的名字，就可以用一个list来表示：



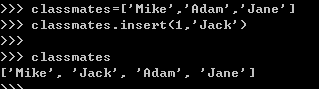
在上面的例子里，变量classmates就是一个list。我们还可以用len()函数来获取list元素的个数。

1509348425(1)

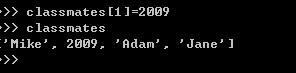
用索引来访问list中每一个位置的元素，记得索引是从0开始的。当索引超出了范围时，Python会报一个IndexError错误，所以，要确保索引不要越界，记得最后一个元素的索引是len(classmates) - 1。如果要取最后一个元素，除了计算索引位置外，还可以用-1做索引，直接获取最后一个元素。



List是一个可变的有序表，所以，可以往list中追加元素到末尾，也可以把元素插入到指定的位置，要删除指定位置的元素，用pop（）方法，



要把某个元素替换成别的元素，可以直接赋值给对应的索引位置，list里面的元素的数据类型也可以不同，比如下图的这个例子。

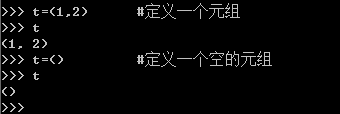


不过因为list里面可以包含不同的数据类型，这也造成它运行起来可能会比较慢，为了弥补这个缺陷，我们将来会引用数组这一个概念。

### 1.3.3数据类型——列表

接下来我们来讨论另一种有序列表，叫做元组：tuple。tuple和list非常类似，但是tuple一旦初始化就不能修改，比如同样是列出同学的名字，现在，classmates这个tuple不能变了，它也没insert()这样的方法。其他获取元素的方法和list是一样的，你可以正常地使用classmates[0]，classmates[-1]，但不能赋值成另外的元素。

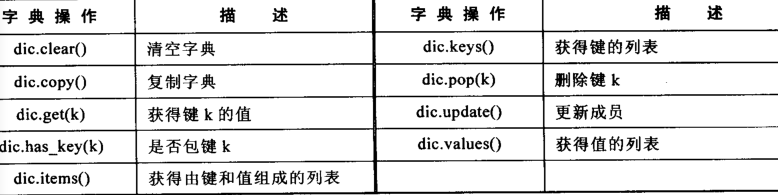
不可变的tuple有什么意义？因为tuple不可变，所以代码更安全。如果可能，能用tuple代替list就尽量用tuple。所以当你定义一个tuple的时候，tuple的元素就必须被确定下来。



需要注意的是，要定义一个只有1个元素的tuple，如果你这么定义：t=(1),定义的不是tuple，是1这个数！这是因为括号()既可以表示tuple，又可以表示数学公式中的小括号，这就产生了歧义，因此，Python规定，这种情况下，按小括号进行计算，计算结果自然是1。所以，只有1个元素的tuple定义时必须加一个逗号，来消除歧义：t=(1,)。而Python在显示只有1个元素的tuple时，也会加一个逗号，以免你误解成数学计算意义上的括号。

### 1.3.4 数据类型——字典

字典是Python中比较特别的一类数据类型，以大括号“{}”包围的数据集合。字典与列表的最大不同在于，字典是无序的，在字典中是通过键来访问成员的。字典也是可变的，可以包含任何其他类型，不过，字典中的成员位置只是象征性的，并不能通过其位置来访问该成员。字典中的成员是以“键：值”的形式来声明的，常用字典操作都在下面这张表里：



### 1.3.5基本语句

语句是Python脚本的基础，它能帮助我们根据一定的条件来执行脚本中的不同语句，从而完成不同的功能。

首先，我们来看条件语句。计算机之所以能做很多自动化的任务，因为它可以自己做条件判断。，在Python程序中，它是if语句是基本的的条件测试语句，用来判断可能遇到的不同情况，以及在不同情况下应该进行的下一步操作。它的基本形式如下：

If <条件>: #设为条件A

<语句> #要用缩进来表示语句处于if语句之中，设为语句A

elif <条件>： #设为条件B

<语句> #设为语句B

else：

<语句> #设为语句C

用简单的话来说就是，首先判断输入的数据是否满足条件A，如果满足就执行条件A，不满足的话就接着判断是否符合条件B的要求，如果满足了就执行语句B，否则就执行语句C。

第二个要介绍的是循环语句。要计算1+2+3，我们可以直接写表达式：

IMG_256

要计算1+2+3+...+10，勉强也能写出来。但是，要计算1+2+3+...+10000，直接写表达式就不可能了。为了让计算机能计算成千上万次的重复运算，我们就需要循环语句。

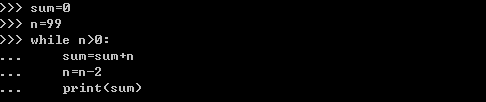
Python的循环有两种，一种是for...in循环，依次把list或tuple中的每个元素迭代出来，如下面的例子所示。

IMG_256

执行这段代码，会依次打印names的每一个元素：

IMG_256

第二种循环是while循环，只要条件满足，就不断循环，条件不满足时退出循环。比如我们要计算100以内所有奇数之和，可以用while循环实现：



上面这段例子就是在循环内部变量n不断自减，直到变为-1时，不再满足while条件，循环退出。

while语句与for语句不同的是，while语句中只有在测试条件为假时才会停止，所以在while的语句块中一定要包含改变条件的语句，以保证循环能够结束，避免死循环的出现。

### 1.3.6小结

Python的语法比较简单，采用缩进方式。

缩进有利有弊。好处是强迫你写出格式化的代码，但没有规定缩进是几个空格还是Tab。按照约定俗成的管理，应该始终坚持使用4个空格的缩进。缩进的另一个好处是强迫你写出缩进较少的代码，你会倾向于把一段很长的代码拆分成若干函数，从而得到缩进较少的代码。

缩进的坏处就是“复制－粘贴”功能失效了，这是麻烦的地方。当你重构代码时，粘贴过去的代码必须重新检查缩进是否正确。此外，IDE很难像格式化Java代码那样格式化Python代码。

最后，请务必注意，Python程序是大小写敏感的，如果写错了大小写，程序会报错。

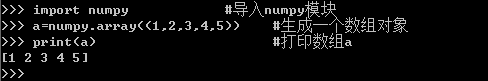
## 1.4第三方包

Python本身就内置了很多非常有用的模块，只要安装完毕，这些模块就可以立刻使用。除了内建的模块外，Python还有大量的第三方模块。基本上，所有的第三方模块都会在PyPI - the Python Package Index上注册，只要找到对应的模块名字，即可用pip安装。而由于前面本书已经推荐大家安装了Anaconda，一些常用第三方库已经包括在里面，省去了我们很多需要自行安装的环节。

科学计算是计算机应用的主要内容之一，也是机器学习的很重要的部分，NumPy和SciPy是Python中用以实现科学计算的模块包，Matplotlib是绘制2D图形的一种模块，Sklearn包含有完善的文档和丰富的机器学习算法，这也是Python语言如此好用的原因之一。接下来我们主要介绍一下这几个第三方包。

### 1.4.1 NumPy简介

NumPy提供了Python没有提供的数组对象，使用了NumPy的数组对象就可以创建类似于C语言中的数组。使用NumPy时要先导入NumPy模块，如下图所示。

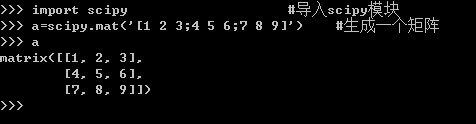


上面这个例子里，

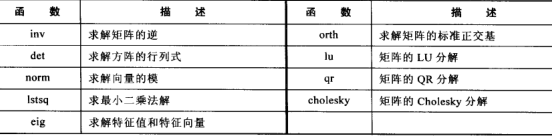
NumPy主要为后面的SciPy和Matplotlib提供了基础的数据结构。

### 1.4.2 SciPy简介

SciPy模块依赖于NumPy，但是SciPy提供了更多的数学工具。使用SciPy不仅可以进行矩阵运算，还可以用来求解线性方程组、进行积分运算、优化等。SciPy的功能非常接近Matlab。SciPy简单的运用如下图所示：



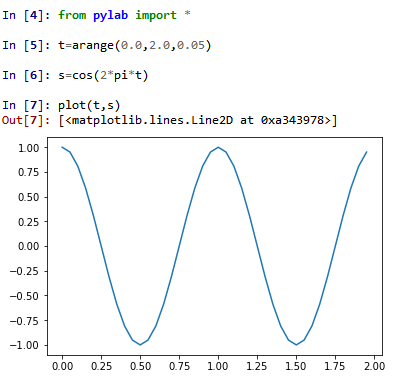
使用SciPy可以完成矩阵分解、线性方程组求解和多项式求根数学运算。SciPy中的函数名与Matlab中的函数名大部分都相同，用法也差不多。熟悉Matlab的用户可以很快的熟悉这个第三方包。下图是一些常用的函数，以后在机器学习中很多也都会被用到：



### 1.4.3 Matplotlib简介

在Python中，我们可以使用Matplotlib来绘制函数图形，它依赖于NumPy与SciPy。和SciPy一样，其中的大部分函数都和Matlab中的函数名相同。它可以绘制多种形式的图形包括普通的线图、直方图、饼图、散点图以及误差线图等。

使用Matplotlib绘制图形主要使用它的plot函数，Matplotlib还包含了一些用于设置x、y轴标签文本以及图形标题的函数。同时它还支持一部分Tex排版命令，可以较好地显示数学公式。接下来我们将在IPython交互式命令行中使用Matplotlib绘制cos图形。



### 1.4.4 scikit learn简介

Sklearn包括了很多种机器学习的类型，比如说监督学习和非监督学习，在监督学习中，又可以分为分类学习和线性回归学习。当我们想训练我们的电脑来识别一些图片，比如猫狗猪牛等等时，这个叫做分类学习。当我们想预测房价高低时，用到的就是线性回归学习。非监督学习就是指，我们不告诉计算机你看到了什么东西，让计算机自己琢磨出来这些数据有什么不同，再画一条线区分开来这些数据的类型。从上面的大概了解可以看出来Sklearn是非常实用的。