

装订线内不要答题

题号	一	二	三	总分	阅卷人
得分					

1. 要求 sed 每隔 2 行输出文件 f 的内容，采用命令\_\_\_\_\_；欲将文件逆序输出（从最后一行到第一行），则命令为\_\_\_\_\_。

2. 如果两个文件有相同的 inode，则修改其中一个文件是否会影响另一个文件？  
\_\_\_\_\_；符号链接的文件与源文件是否有相同的 inode？\_\_\_\_\_。

3. 要实现二维向量  $x$  旋转  $\theta$  角度且长度不变, 需进行正交变换, 也就是用二阶正交矩阵  $A$  与其相乘,  $A =$  。

4. Python 脚本中 `sys.argv[0]` 表示\_\_\_\_\_，而 Bash 函数中的 `$@` 表示\_\_\_\_\_。

5. Linux 下的标准输入输出流包括\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。

6. 对于字符串变量 a, `#{a}` 返回的是\_\_\_\_\_；而如果 a 为数组变量，返回的则是\_\_\_\_\_。

7. 命令 `declare -a` 的目的是\_\_\_\_\_，而 `declare -i` 则\_\_\_\_\_。

8.  $\{x:-5\}$ 、 $\{x:=5\}$ 、 $\{x:+5\}$  分别实现的功能为\_\_\_\_\_、\_\_\_\_\_和\_\_\_\_\_。

9. 栈是一种\_\_\_\_\_的数据结构，而队列相反，是一种\_\_\_\_\_的数据结构。

10. 输出长度为  $n(n > 8)$  的单向链表的第 8 个元素和倒数第 8 个元素的时间复杂度分别为 \_\_\_\_\_ 和 \_\_\_\_\_。

1. 下面哪些选项可以输出文件的第 99 行内容？

B. sed '99p' file

D. head -99 file | tail -1

2. 关于命令 `chmod` 文件权限分配正确的是？

- A. `s` 分配的是 SUID 的权限
- B. `t` 分配的是 sticky bit 的权限
- C. `chmod 777` 是一种非安全行为
- D. `chmod 4770` 分配的是 SGID 权限

3. 用 `ls -l` 查看文件信息的时候，哪些是正确的？

- A. 硬连接文件的类型位是 `l`
- B. 目录文件的类型位是 `d`
- C. 常规文件的类型位是 `r`
- D. 字符设备文件的类型位是 `b`

4. 下面哪些命令可用来比较文件内容？

- A. `comm`
- B. `diff`
- C. `cat`
- D. `tac`

5. 关于正则表达式中问号的作用，正确的是？

- A. `xx?` 表示字母 `x` 的 1-2 次的重复
- B. `x??` 是非贪婪的匹配方式，仅匹配一次 `x`
- C. 问号可用于零宽断言
- D. 问号可用于分支匹配

6. 关于 Python 的 `sys` 模块，正确的是？

- A. `sys` 模块包含了操作系统的配置
- B. `sys` 模块包含的是 Python 解释器的配置和相关函数
- C. `sys.id()` 输出的是目标对象的内存地址
- D. `sys.getsizeof()` 输出的是目标对象的被引用数

7. 下面哪个不是 Python 的内置函数？

- A. `len`
- B. `size`
- C. `id`
- D. `getattr`

8. 下面哪个表示的是 `bash` 脚本的第 11 个位置参数？

- A. `$11`
- B. `${11}`
- C. `${@:11:1}`
- D. `${*[11]}`

9. 关于 python 函数的位置参数和关键词参数，正确的是？

- A. python 一直这两者参数
- B. 当输入 `list` 作为参数时，当作位置参数
- C. 当输入 `dict` 作为参数时，作为关键词参数
- D. 参数的个数必须是固定的

10. 关于 python 的常用数据结构，错误的是？

- A. `None` 是一种特殊的数值类型
- B. `tuple` 是不可迭代的数据结构
- C. `tuple` 可作为 `dict` 的 `key`
- D. `list` 是可变的数据结构

11. 有关 NumPy，错误的是？

- A. `numpy.random` 可用于数值模拟
- B. `numpy.linalg` 中包含了大部分的线性代数的运算，如 `svd`

C. numpy 中的简单数值运算与 math 模块中的运算完全相同, 例如对数 `numpy.log==math.log`

D. `numpy.random` 可以计算 Gaussian 分布的 pdf

12. 关于常用的数据结构, 错误的是?

A. 数组、链表、队列、栈都是线性的数据结构

B. 图和树都是网状的数据结构

C. 图和树都只能用数组实现

D. 线性数据结构都只有唯一的遍历方式, 而树有三种遍历方式

13. 下面是单向链表的遍历实现, 哪些说法是错误的?

```
def traverse(self):  
    temp = self.head  
    while temp.next is not None:  
        print temp.data  
        temp = temp.next
```

A. 当链表为空时会抛出 `AttributeError` 错误 B. 当链表不为空时不能输出链表的 tail

C. 程序错误因为不能一直修改 temp D. 只知道链表的 head 无法实现链表的遍历

14. 如果一个单向链表同时知道其 head 和 tail, 则下面哪些对链表的操作与链表长度无关?

A. 删除链表的第一个元素

B. 在链表头部插入一个元素

C. 删除链表的最后一个元素

D. 在链表的最后添加一个元素

15. 下面的函数以单向链表 1->2->3->4->5->6 的 head 为参数, 则输出结果为?

```
def fun(head):  
    if start is None:  
        return  
    print start.data,  
  
    if start.next is not None:
```

```
        fun(start.next.next)

    print start.data,
```

A. 1 4 6 6 4 1      B. 1 3 5 1 3 5      C. 1 2 3 5      D. 1 3 5 5 3 1

16. 如果想要反转一个单向链表，那么需要在'''ADD A STATEMENT HERE'''处填入哪个？

```
def reverse(head):
    prev = None
    current = head
    while current is not None:
        next = current.next
        current.next = prev
        prev = current
        current = next
    }
    '''ADD A STATEMENT HERE'''
}
```

A. head=prev      B. head=current      C. head=next      D. head=None

17. 关于 SVM，错误的有？

- A. SVM 是一类间隔最大化方法
- B. 核函数计算的是将特征投影到目标高维空间后特征向量之间的内积
- C. 核函数仅针对线性不可分问题
- D. SVM 可直接对原问题求解，而必须对对偶问题求解

18. 关于零宽断言 (zero-length assertion)，正确的有？

- A. (?=exp) 为正向先行断言，表示该位置的后面必须为表达式 exp
- B. (?<!exp) 为负向先行断言，表示该位置的后面不能是表达式 exp
- C. (?=exp) 为正向后发断言，表示该位置的前面必须为表达式 exp
- D. (?!exp) 为负向后发断言，表示该位置的前面不能是表达式 exp

19. 关于正则表达式，错误的是？

- A. 正则表达式一般分为 BRE, ERE 和 PRE B. grep 欲要支持 ERE, 必须使用选项 -P  
C. sed 欲要支持 ERE, 必须使用选项 -E D. Python 的 re 模块默认支持所有的 ERE

20. Logistic 函数  $y = \frac{e^{\beta^T x}}{e^{\beta^T x} + 1}$  的一阶导数是？

- A.  $y(1 - y)$  B.  $y(1 + y)$  C.  $y$  D.  $1 - y$

三、解答题：（任选 10 小题, 每小题 4 分, 共 40 分）

1. 创建一个目录, 使得同组的任意其他用户可以创建和访问该目录, 但不能访问、修改和删除该目录下其他同组用户创建的文件, 该怎么做？

2. 用 awk 读入一个文件, 并统计文件中每个单词出现的频率, 并输出使用频率最高的前 10 个单词。

3. 写一个 bash 函数, 输出 Fibonacci 数列的前 10 个数。

4. “幸运数字”指的是其每个数值连续相加后得到 7 的数字, 例如 62431, 其和为  $6+2+3+4+1=16$ , 而  $1+6=7$ , 故该数字为幸运数字。现在需要你输出从 100 到 10000 之间的所有幸运数字。

5. 将一个 7 位数的数值倒序输出, 注意的是, 需要将前面的零给去掉。分别用 Python 和 Bash 实现。

6. 我们在课上提到过, 在定义 Python 类的时候, 函数名前面加上双下划线的方法是特殊方法, 请举例说明有哪些常见的特殊方法。

7. 实现一个 Python 类 NumList, 作为 list 类的派生类, 要求其可以实现下列功能:

(1) 每个元素必须为数值, 可以是 int, 也可以是 float;

(2) 实现与一个标量的乘法和除法, 与 list 不同的是, 要求乘法计算的结果是其每个元素的乘积;

(3) 实现与另一个 NumList 的加法和乘法, 可以元素个数不同。

这是例子:

```
a = NumList([1,2])
b = NumList([2,3,4,5,6])
print a*3 # 输出NumList([3,6])
```

```
print a/2 # 输出NumList([0.5,1])
print 3*a # 输出NumList([3,6])
print a*b # 输出NumList([2,6,4,10,6])
print a+b # 输出NumList([3,5,5,7,7])
```

8. 说说最小二乘法的原理，探讨一下其在 linear regression 中的应用。
9. 下面是一组数据，分别给出了 6 个研究对象的身高、体重、BMI 和体重分级。请用 kNN 算法，估计一个新的研究对象（身高 165，体重 60）的 BMI 以及体重分级，设  $k = 3$ 。请问这里使用的单位是否会影响我们的判断？

	身高	体重	BMI	体重分级
Person1	180	85	24	正常
Person2	185	70	19	偏瘦
Person3	156	40	17	偏瘦
Person4	190	150	35	偏胖
Person5	189	78	23	正常
Person6	200	160	38	偏胖

10. pandas 模块有哪几种数据结构，与内置数据结构相比，有什么优势？
11. 梯度下降法和牛顿法在求解最优化问题时各有什么样的优缺点？请举例说明。
12. 用 numpy.random 模块结合 matplotlib.pyplot.hist() 验证概率论中的大数定理。