

苏州大学

2010年攻读硕士学位研究生入学考试试题

专业名称: 系统生物学

考试科目: 专业综合2 (A)卷

生物信息学基础、Perl 编程基础、线性代数三组试题中任选一组, 多选作废

注意: 答案请不要做在试题纸上。

I、 生物信息学基础 (选这一组的不要做 Perl 编程基础、线性代数)

一. 生物信息学软件说明 (每题 8 分, 共 80 分)

1. Tree View
2. DnaSP
3. SEAVIEW
4. SOSUI
5. ESPript
6. Seq-Gen
7. FatGO
8. fastDNAmI
9. MatInspector
10. InterProScan

二. 名词解释 (每题 10 分, 共 50 分)

1. Profile hidden Markov models
2. PAM matrix
3. statistical significance
4. Systems biology markup language (SBML)
5. Molecular clock

三. 问答题 (共 20 分)

1. What kind of model could be used to detect remote protein homologies?

II、 Perl 编程基础 (选这一组的不要做生物信息学基础、线性代数)

(一) 判断对错。如果有错, 请解释理由。(3 x 10=30 分)

- a) 调用 print 函数时, 肯定从一个新行的起始处开始打印。
- b) Perl 认为 "number" 和 "NuMbEr" 这两个字串是完全一致的。
- c) *, /, %, + 和 - 这几个算术运算符的运算顺序是相同的。
- d) 假如在数量上, 初始化列表中的值要少于数组中的元素, 那么剩下那些无值可赋的元素会自动初始化为 undef。
- e) 如将一个数组作为双引号字串的一部分输出, 数组内所有元素会合并显示成一个长字串, 元素间没有任何空格。
- f) Perl 会把最近出现的系统错误提示信息保存在变量 \$? 中。
- g) 除 do/while 和 do/until 结构之外的所有循环块都可使用一个标签。
- h) 只有 \$x > \$y 和 \$w < \$z 同时为 "真", 条件表达式 (\$x > \$y || \$w < \$z) 才会为真。
- i) 特殊变量 @_ 会展开所有数组, 但却不能展开散列。
- j) 每生成一个新包, Perl 都会创建一个新的、空白的符号表。

(二) 指出下述语句的错误并改正: (4 x 5=20 分)

- k) if (\$c=7) {print "c is equal to 7\n";}
- l) if (\$c<7); {print "c is less than 7\n";}
- m) @array[3]; #obtain the value at location 3
- n) \$string =~ m/\${d*}^/;
- o) \$string =~ s/(?=<i>be/am/;

(三) 请写出一些语句, 完成下述要求: (4 x 5=20 分)

- p) 创建一个名为 numbers 的数组, 把它的元素设为 1 到 100 的整数。
- q) 把一个包含了 0 值的新元素添加到 numbers 数组的起始处。
- r) 打印 numbers 数组, 并用空格分隔每个元素。
- s) 用一个 for 循环输出每个元素。
- t) 使用列表赋值技术, 交换 numbers 数组位置 0 和 1 处的值。
- u) 使用 splice 函数, 删除数组的头两个元素。

(四) 写一个程序, 完成下述任务: (10 分)

- v) 创建一个名为 @array 的数组, 在其中包含 1~99 之间的整数。(2 分)
- w) 使用 grep 函数创建一个名为 @oddArray 的新数组, 其中只包括 @array 中值为奇数的元素。(4 分)
- x) 使用 foreach 结构, 求 @oddArray 中的整数总和。(4 分)
- y) 打印出结果。(2 分)

(五) 写一个子程序, 名为 `above_average`, 将一系列的数字作为其参数, 返回所有大于平均值的数字。利用下面的程序进行测试: (10分)

```
my @fred=&above_average(1..10);
print "\@fred is @fred\n";
print "(Should be 6 7 8 9 10)\n";
my @barney=&above_average(100,1..10);
print "\@barney is @barney\n";
print "(Should be just 100)\n";
```

(六) 假设有一字符串保存在标量变量 `$string` 中, 用正则表达式分别统计字符串中的数字、字母、空白字符和单词的个数。(3+3+3+6=15分)

(七) 写一个程序, 它可以得到当前文本文件的一个拷贝。在拷贝的文件中, 字符串 `Fred` (大小写无关) 将被 `Larry` 替换掉。(因此, "`ManfredMann`" 将变成 "`ManLarryMann`") 输入的文件名已经在命令行中指定 (不需要询问用户), 输出的文件名是对应的输入文件名后面加上 `.out`。(15分)

(八) 写一个程序, 找出命令中存在时间最长的文件名, 并报告其天数。注意处理命令行参数为空的情况。(15分)

(九) 写一个程序, 从文件中读入字符串, 一行一个字符串, 然后让用户输入模式, 这个模式可能匹配上某些字符串。对于每一个模式, 程序将指出文件中有多少个字符串 (多少行) 匹配上了, 并指出是哪些。对于新的模式不需要重新读文件, 将这些字符串保留在内存中。文件名可以直接写在程序之中。如果模式无效 (例如, 圆括号不匹配), 则程序报告这个错误, 并让用户继续尝试新的模式。当用户输入一个空行, 则程序退出。(15分)

III、 线性代数 (选这一组的不要做生物信息学基础、Perl 编程基础)

一, 对矩阵 $A = \begin{pmatrix} 2 & 2 & 3 \\ 4 & 7 & 7 \\ -2 & 4 & 5 \end{pmatrix}$

- (1) 求矩阵 A 的迹, 行列式和逆矩阵; (15分)
- (2) 求解线性方程组: $Ax = b$, 其中 $b = (7, 18, 7)^T$ 。(10分)

二, n 阶矩阵 $A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & \cdots & \frac{1}{2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdots & 1 \end{pmatrix}_{n \times n}$

- (1) 求 A 的行列式; (15分)
- (2) 判断 A 是否是正定矩阵, 并说明理由。(10分)

三, 对矩阵 $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & -1 \\ 1 & 0 & 1 \end{pmatrix}$

- (1) 求 A 所有的特征值和特征向量; (10分)
- (2) 将矩阵 A 对角化, 即求可逆矩阵 P , 使得 $P^{-1}AP$ 为对角矩阵; (10分)
- (3) 计算 A^{10} 。(5分)

四, A 是正交矩阵, 证明
 (1) A 的行列式等于 1 或 -1; (5分)
 (2) A 的特征值等于 1 或 -1。(10分)

五, 矩阵 $A = (a_{ij})_{n \times n}$ 是非奇异的下三角矩阵, 即 $a_{ij} = 0$ 当 $j > i$ 且 $|A| \neq 0$, 证明 A 的逆矩阵仍然是下三角矩阵。(15分)

六 矩阵 A 是 $m \times n$ 的实矩阵, 证明
 (1) $A^T A$ 是对称半正定矩阵; (10分)
 (2) 当 A 列满秩时, $A^T A$ 是对称正定矩阵。(15分)

七, A, B 分别为 $m \times n$ 和 $n \times m$ 的实矩阵, 证明 AB 和 BA 的非零特征值相同 (20分)