1.[单选题]

主机甲与主机乙之间已建立一个TCP连接，主机甲向主机乙发送了两个连续的TCP段，分别包含300B和500B的有效载荷，第一个段的序列号为200，主机乙正确接收到这两个数据段后，发送给主机甲的确认序列号是

200

500

800

1000

2. [单选题]

在支持多线程的系统中，进程P创建的若干个线程不能共享的是

进程P的代码段

进程P中打开的文件

进程P的全局变量

进程P中某线程的栈指针

3.[单选题]

排序算法的效率取决于元素的比较次数与元素的位置移动次数，现需要对数组进行升序排序，已知一数组的元素为{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}，问下面哪种排序算法的效率最高？

插入排序

选择排序

快速排序

冒泡排序

4. [单选题]

一个栈的入栈序列是a b c d e，则栈的输出序列不可能是

dceab

decba

edcba

abcde

5. [不定项选择题]

对数据库，关于索引的理解正确的是

创建索引能提高数据插入的性能

索引应该根据具体的检索需求来创建，在选择性好的列上创建索引

索引并非越多越好

建立索引可加速查询

6. [不定项选择题]

用浏览器访问一个Internet网站，可能使用到的协议有

PPP

HTTP

POP

ARP

7. [不定项选择题]

查找或删除性能较低的数据结构有

有序数组

有序链表

AVL树

Hash表

8. [不定项选择题]

以下哪些与编译器的任务有关？

公共子表达式合并

运行程序前加载其依赖的动态库

尾递归优化

常量、不变式预计算

9[填空题]

现有代码如下，则 func(5)的返回值为 1

int func(int n){

if(n <= 1){

return 1;

}else{

return n \* func(n-1);

}

}

10. [填空题]

下面C程序的运行打印结果是\_\_\_\_\_\_\_

#include <stdio.h>

int main(int argc, char\*\* argv) {

char\* array[] = {"hello", "my", "world", "goodbye"};

char\*\* p = array;

p = p + 2;

printf("%s", \*p);

return 0;

}

11. [填空题]

一个长度为100的循环链表，指针A和指针B都指向了链表中的同一个节点，A以步长为1向前移动，B以步长为3向前移动，最少需要同时移动\_\_\_\_\_ 步A和B才能再次指向同一个节点

12. [填空题]

一棵完全二叉树中有33个结点，则该完全二叉树的深度为\_\_\_\_\_\_\_

13.[填空题]

{0, 2, 1, 4, 3, 9, 5, 8, 6, 7}是以数组形式存储的最小堆，删除堆顶元素0后的堆的新结果是\_\_\_\_\_\_\_（结果需要英文逗号分隔）

14. [填空题]

已知关键字序列为(51,22,83,46,75,18,68,30),进行快速排序，第一趟按关键码字51进行,完成后的序列为\_\_\_\_\_\_\_\_\_\_（结果需要英文逗号分隔）

15. [填空题]

如果下列的公式成立：77+77=121，则数字是采用 \_\_\_\_\_进制表示的

16. [填空题]

LRU的cache长度为3,初始为空。依次访问如下元素后，cache里的内容是  
\_\_\_\_\_\_\_\_\_（结果需要英文逗号分隔）

A,A,B,C,A,D,C,E

17. [填空题]

已知一算数表达式的中缀表达式为 a-(b+c/d)\*e，其后缀形式为\_\_\_\_\_

18. [填空题]

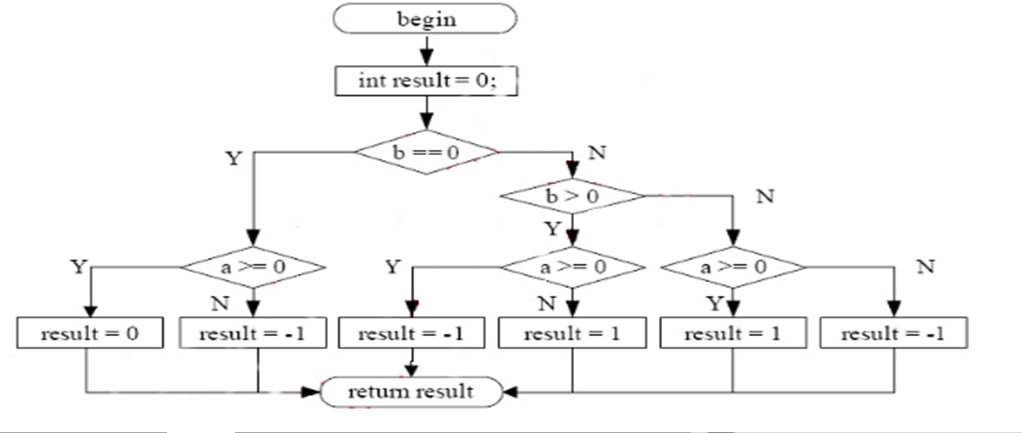
从1 - 1001中，能被数字2或者数字3或者数字5整除的数字有 \_\_\_\_\_个

19. [填空题]

5个盒子每个里面各有一个球，把球全拿出来打乱再放回去，每个球都不在自己原来的盒子里，有\_\_\_\_\_\_\_种可能

20. [填空题]

【T】当用分支覆盖法对以下流程图进行测试时，至少需要设计\_\_\_\_\_个测试用例



21. [问答题]

题目描述

对一个整数的四则运算后缀表达式，请写函数将其打印成日常我们使用的中缀表达式。如对ab+c\*，打印出 (a+b)\*c 。后缀表达式以一个列表形式作为函数输入，列表的元素为数字或加减乘除操作符。

22. [问答题]

题目描述

挖雷游戏是一个N\*N格子棋盘，一些随机的格子里有雷，把所有不是雷的格子挖开游戏取胜结束，挖中了任一有雷的格子游戏失败结束。在挖开一个没有雷的格子时，格子上会显示数字，表示相邻的8个格子里有几颗雷，如果是0颗，则程序会帮助把相邻的格子自动全挖开，如果其中又有0颗的，则继续下去。

请你写函数实现对挖开一个0颗雷的格子后，程序自动处理的过程。布了雷的所有格子的坐标作为已知的输入条件。

23. [问答题]

题目描述

考虑一个网络服务，希望具备防刷的安全特性。假设要求策略是对每次请求访问，如果该请求的来源IP，在当前的前N秒内已经请求过了M次，则拒绝服务X秒。请设计方案，无需写出完全代码，描述清楚设计实现即可。并请针对设计出的方案分析利弊

24. [问答题]

题目描述

（数据结构与算法）列举至少2种排序算法（如快排），并写出实现代码

25. 问答题]

题目描述

（数据结构与算法）已知一随机发生器，产生0的概率是P，产生1的概率是1-P。现在需要构造一个发生器，使得它构造0和1的概率均为1/2，请写出思路或伪代码

26. [问答题]

题目描述

（机器学习理论）请列举生成模型与判别模型的区别

27.[问答题]

题目描述

（机器学习理论）请列举分类模型和回归模型的区别

28. [问答题]

题目描述

（机器学习理论）请描述决策树的原理、过程、终止条件，以及如何防止过拟合

29.[问答题]

题目描述

（机器学习理论）请描述K-means的原理，说明选择聚类中心的方法

30. [问答题]

题目描述

（机器学习理论）请描述推荐系统中协同过滤算法的原理

31. [问答题]

题目描述

（NLP方向）什么是中文分词？ 列举出几种你所知的分词方法

32. [问答题]

题目描述

（NLP方向）现在深度学习在nlp领域有哪些应用？ 请具体说明

33. [问答题]

题目描述

（图像方向）传统图像处理提取的sift特征是什么意思？

34. [问答题]

题目描述

（图像方向）什么是卷积神经网络？请说明卷积的意义

35. [问答题]

题目描述

（深度学习方向）深度学习和过往的神经网络训练方法有什么区别？

36. [问答题]

题目描述

（深度学习方向）深度学习和过往的神经网络训练方法有什么区别？列举几种深度学习的loss function，并说明意义

37. [填空题]

某校计算机专业有80%的学生参加过比赛A，有30%的学生参加过比赛B，有15%的学生两个比赛都没参加过，随机询问一个学生，该学生既参加过比赛A又参加过比赛B的概率为\_\_\_\_\_

38. [填空题]

某种酒精检测仪在对吸烟的人使用时98%报阳性、2%报阴性，而在对不吸烟的人使用时98%报阴性、2%报阳性。已知某校吸烟的学生大概占2%，则若对该校某学生的检验结果呈阳性，那么该学生吸烟的概率为 \_\_\_\_\_\_，计算公式为 \_\_\_\_\_\_\_

39. [填空题]

考虑值集{6,47,49,15,42,41,7,39,43,40,36}，其四分位数极差为\_\_\_\_\_\_\_\_\_\_

40. [填空题]

从某超市顾客中随机抽取5名，他们的购物篮数据的二元0/1表示如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 面包 | 牛奶 | 尿布 | 啤酒 | 鸡蛋 | 可乐 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 1 | 0 | 1 | 1 | 1 | 0 |
| 3 | 0 | 1 | 1 | 1 | 0 | 1 |
| 4 | 1 | 1 | 1 | 1 | 0 | 0 |
| 5 | 1 | 1 | 1 | 0 | 0 | 1 |

某学生依据此数据做关联分析，考虑规则{牛奶，尿布}->{啤酒}，则该规则的支持度（support）为\_\_\_\_\_\_，置信度（confidence）为\_\_\_\_\_\_\_\_

41. [问答题]

题目描述

列举至少2种排序算法（如快排），并写出实现代码

42. [问答题]

题目描述

什么是欠拟合、过拟合？避免过拟合有哪些途径？

43. [问答题]

题目描述

请描述K-means的原理，说明选择聚类中心的方法？

44. [问答题]

题目描述

梯度下降法求解最优化问题的原理与步骤？

45. [问答题]

题目描述

请描述推荐系统中协同过滤算法的原理？

46. [问答题]

题目描述

请说明分类算法中常用的评估指标（准确率、召回率、ROC、AUC）的定义，及计算方法？

1.D

2.D

3.A

4.A

5.BCD

6.ABD

7.AB

8.ACD

9.120

10.world

11.50

12.6

13. {1,2,5,4,3,9,7,8,6} 或 1,2,5,4,3,9,7,8,6

14.30,22,18,46,51,75,68,83 或 （30,22,18,46,51,75,68,83）

15.13

16.E,C,D

17.abcd/+e\*-

18.734

19.44

20.6

21.参考要点：使用一个栈，扫描输入列表入栈，碰到操作符时，将前两个元素出栈，与操作符组成中缀子表达式的字符串再入栈。组成字符串时，要考虑操作符两边的元素是否要加括号。扫描完后栈中应剩下唯一字符串，打印即可。注意输入表达式的错误处理。

示意伪代码：

function printInfixExpression(list postfixExpression)

{ Stack stack;

postfixExpression.foreach( x=> {

if(x is Number) { stack.push(x); }

else if(x is Operator) {

if(stack.size() < 2) { printErrorThenReturn(); }

left = stack.pop(); right = stack.pop();

if (left is ExpressionString and left.operator.priority < x.priority) {

left.text = ‘(‘ + left.text + ‘)’;

}

if (right is ExpressionString and right.operator.priority <= x.priority) { //这里操作符优先级判断要用小于等于

right.text = ‘(‘ + right.text + ‘)’; }

}

subExpr = new ExpressionString;

subExpr.operator = x;

subExpr.text = stringAppend(left, x, right);

stack.push(subExpr);

}

// 既不是数字也不是操作符，输入错误

else { printErrorThenReturn(); }

} // end of lambda

); // end of foreach

if(stack.size() > 1){ // 输入表达式有错误

printErrorThenReturn();

}

print(stack.pop()O); // 栈中剩下唯一元素，为结果表达式字符串。

}

22.参考要点：自动挖开的所有0颗雷的格子是一个连通图（每个节点最多4条边），函数通过计算当前节点周围的雷数量来不断发现相邻节点，边发现边实现图的遍历，遍历过程就是将相应格子的状态从“未挖开”变为“挖开“。

代码容易出现的错误：

遍历时未考虑回到原点形成无限循环的状况，遍历时未考虑雷区N\*N的边界情况

23.参考要点：

每个IP关联一个最近访问时间戳T，和一个链表L，链表元素为一个细粒度时间段内的访问次数，比如粒度定为500毫秒，则链表内元素依次代表了该IP当前500毫秒的（开始时间，结束时间，请求次数），往前的500毫秒中（开始时间，结束时间，请求次数），再往前500毫秒的（开始时间，结束时间，请求次数）。。。相邻元素的时间段不一定需要连续。

用hash关联IP保存所有T和L

对每一个来访请求，按IP从hash中找到关联的链表L，将最新的元素计数加1（根据当前时间决定是否创建新元素），将N秒前的元素删除，如果剩余元素的计数之和大于等于M，并且最近访问时间戳T到当前未超过X秒，则拒绝服务，否则更新T并提供服务。

定时器将过期IP数据从hash中删除

处理过程中使用同步结构，保证并发安全。

24.

25.

26.

27.

28.

29.

30.

31.

32.

33.

34.

35.

36.

37.25%

38.(1) 50%

(2) 0.98\*0.02/(0.98\*0.02+0.02\*0.98)

39.28

40.(1) 0.4 或 0.40

(2) 0.67

41.

42.

43.

44.

45.

46.