

一、单选题

1. 你有 1000 枚 1 元硬币,999 个都是正常的,只有 1 个硬币正反面都是国徽。假设你随机抽了 1 枚硬币出来,你抛这枚硬币 10 次,每一次都是国徽朝上。请问你抽中的是那枚两面国徽的硬币的概率是多少?

A:≈99.9%

B:≈99%

C:≈50%

D:≈1%

E:≈0.1%

2. 假设你是本市排名第一的富商,某天你被 2 个海盗绑架到一条船上,海盗们拿到了赎金,但幸运的是这 2 个海盗因为赎金分配大打出手,而你挣脱了绳索并抢到了一把枪。现在你们 3 人每人都有把枪,互相举枪指向别人。你们 3 人只有 1 人能够幸存。也就是说,如果你要活命,那么这 2 个海盗必须死。假设,你枪法不准,只有 10%的命中率,海盗 A 有 30%的命中率,海盗 B 有 60%的命中率。再假设,你们的***顺序是固定的,你先***,接着是海盗 A,再接着是海盗 B,再接着是你,...,一直轮流直到最后一人存活,而你们的弹药是充足的。请问,你第一枪应该先打谁?

A:海盗 A

B:海盗 B

C:都不打,开空枪

3. Nave Bayes 是一种特殊的 Bayes 分类器特征变量是 X 类别标签是 C,它的一个假定是?

A:各类别的先验概率 $P(C)$ 是相等的

B:以 0 为均值, $\sqrt{2}/2$ 为标准差的正态分布

C:特征变量 X 的各个维度是类别条件独立随机变量

D: $P(X|C)$ 是高斯分布

4. 关于 logit 回归和 SVM 不正确的是

A:Logit 回归目标函数是最小化后验概率

B:Logit 回归可以用于预测事件发生概率的大小

C:SVM 目标是结构风险最小化

D:SVM 可以有效避免模型过拟合

5. 以下几种模型方法属于判别式模型(Discriminative model)的是

A:混合高斯模型

B:条件随机场模型

C:朴素贝叶斯

D:隐马尔科夫模型

二、多项选择题

1. 对给定的关键字序列 110,119,007,911,114,120,122 进行基数排序,则第 2 趟分配收集后得到的关键字序列是 ()

A:7.1101191149111204E18

B:7.1101191149111204E18

C:7.1109111141191199E18

D:1.10120911122114E20

2. 设在内存中有 P1、P2、P3 三道程序,并按照 P1、P2、P3 的优先次序运行,其内部计算和 O 操作时间由下图给出:

P1: 计算 60ms-----I/O 80ms-----计算 20ms

P2: 计算 120ms-----I/O 40ms-----计算 40ms

P3: 计算 40ms-----I/O 80ms-----计算 40ms

调度程序的执行时间忽略不计,完成这三道程序比单道运行节省的时间是 ()

A:80ms

B:120ms

C:160ms

D:200ms

3.关于滑动窗口协议(Sliding Window Protocol),以下说法错误的是:

A:滑动窗口协议工作在网络层

B:滑动窗口协议是为了避免网络拥塞

C:滑动窗口协议可以保证网络包保持顺序、不出错误地提交给应用层

D:当接受方接收到的网络包超过接受窗口时,会将该网络包丢弃

三、问答题

1.对知乎来说,我们对系统进行优化,一方面是为了吸引读者,另一方面也是为了产出高质量、受欢迎的内容。因此,我们需要设计一个系统来预测某一用户撰写某一问题答案的可能性,它将问题通过提示发送给潜在的优质回答者。

请设计一个系统,来实现上述功能(预测某一用户撰写某一问题答案的可能性),简要概述系统架构和基本步骤。

2.一个长度为 $L(L \geq 1)$ 的升序序列 S ,处在第 $L/2$ 个位置的数称为 S 的中位数。例如,若序列 $S_1=(11,13,15,17,19)$,则 S_1 的中位数是 15,两个序列的中位数是含它们所有元素的升序序列的中位数。例如,若 $S_2=(2,4,6,8,20)$,则 S_1 和 S_2 的中位数是 11。现在有两个等长升序序列 A 和 B ,试设计一个在时间和空间两方面都尽可能高效的算法,找出两个序列 A 和 B 的中位数。

要求

(1)给出算法的基本设计思想。

(2)根据设计思想,采用 C 或 C++或 JAVA 语言描述算法,关键之处给出注释。

(3)说明你所设计算法的时间复杂度和空间复杂度。

3.分类问题的评价标准有哪些?ROC 和 AUC 分别是什么,坐标的意义是?

四、编程题

1.题目名称:进制转换

时间限制:3000ms

题目描述:将一个长度最多为 30 位数字的十进制非负整数转换为二进制数输出。

输入描述:多组数据,每行为一个长度不超过 30 位的十进制非负整数。(注意是 10 进制数字的个数可能有 30 个,而非 30bits 的整数)

输出描述:每行输出对应的二进制数。

例

输入 3

输出 11

2.题目名称:目录结构

时间限制:3000ms

题目描述:给你一串路径,譬如:a\b\c a\d\e b\cst d\你把这些路径中蕴含的目录结构给画出来,子目录直接列在父目录下面,并比父目录向右缩一格,就像这样: a b c d e b cst d

同一级的需要按字母顺序排列,不能乱。

输入描述:每个测试案例第一行为一个正整数 $n(n \leq 10)$ 表示有 n 个路径,当 n 为 0 时,测试结束,接下来有 n 行,每行有一个字符串表示一个路径,长度小于 50。

输出描述:输出目录结构,每一个测试样例的输出紧跟一个空行。

示例 1:

输入

4

a\b\c

a\d\e

b\cst

d\

输出

a

b

c

d

e

b

cst

d