

北京大学信息学院考试试卷

考试科目：_____ 姓名：_____ 学号：_____.
 考试时间： 20** 年__月__日 任课教师：

题号	一	二	三	四	五	六	七	八	总分
分数									
阅卷人									

北京大学考场纪律

- 1、考生进入考场后，按照监考老师安排隔位就座，将学生证放在桌面上。无学生证者不能参加考试；迟到超过 15 分钟不得入场。在考试开始 30 分钟后方可交卷出场。
 - 2、除必要的文具和主考教师允许的工具书、参考书、计算器以外，其它所有物品（包括空白纸张、手机、或有存储、编程、查询功能的电子用品等）不得带入座位，已经带入考场的必须放在监考人员指定的位置。
 - 3、考试使用的试题、答卷、草稿纸由监考人员统一发放，考试结束时收回，一律不准带出考场。若有试题印制问题请向监考教师提出，不得向其他考生询问。提前答完试卷，应举手示意请监考人员收卷后方可离开；交卷后不得在考场内逗留或在附近高声交谈。未交卷擅自离开考场，不得重新进入考场答卷。考试结束时间到，考生立即停止答卷，在座位上等待监考人员收卷清点后，方可离场。
 - 4、考生要严格遵守考场规则，在规定时间内独立完成答卷。不准交头接耳，不准偷看、夹带、抄袭或者有意让他人抄袭答题内容，不准接传答案或者试卷等。凡有违纪作弊者，一经发现，当场取消其考试资格，并根据《北京大学本科考试工作与学术规范条例》及相关规定严肃处理。
 - 5、考生须确认自己填写的个人信息真实、准确，并承担信息填写错误带来的一切责任与后果。
- 学校倡议所有考生以北京大学学生的荣誉与诚信答卷，共同维护北京大学的学术声

以下为试题和答题纸，共 14 页。

多重选择题：（每小题 2 分，共 20 分）

从候选答案中选择 1 到多个正确答案。题目中明确单选的选择一个正确答案。

关于 LZW，Run-length 和 Huffman 编码这些压缩算法下面的说法哪些是正确的：

- A) Huffman 编码从信息论的角度来看是最优的因此在所有情况下压缩比也最高，但算法效率相对较低。
- B) LZW 由于生成的词典需要包含在数据包中，因此适用于词条数量小且重复率高的情况。
- C) Run-length 算法适合对低频分量较高的数据流进行压缩。
- D) 如果能对需要压缩的数据找到一个生成函数，则上述三种压缩算法都是浮云。

答案：（ C D ）

5、在图像处理中常常会采用 RGB 三子节色彩编码来表示一个像素的颜色。假定目前一个像素的色彩编码为 FF0000H（16 进制），则该像素的颜色应该为

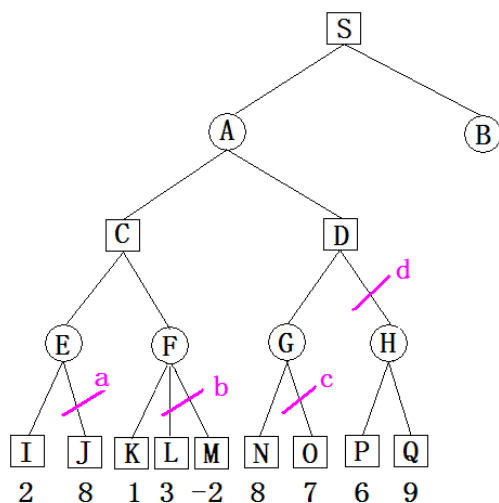
- A. 红色
- B. 蓝色
- C. 灰色
- D. 绿色

答案：（ A ）

6

10. 在下面的博弈树中，如果按照从左向右的正常搜索顺序，哪一支最先会发生 alpha-beta 剪枝。（方框代表甲选手，圆代表乙选手，最下面的得分是指甲选手在该局面下的估分）

- A. a;
- B. b;
- C. c;
- D. d;



答案： B

一、阅读程序并给出运行结果（共 36 分）

请写出下面python程序的运行结果，每题3分。

```
1、
lst = ['a', 'b', ['c', 'd']]
lst2 = lst[:]
lst2[0] = 'e'
lst2[2][0] = 'f'
lst2.append('g')
lst2[2].append('h')
print(lst)
print(lst2)
运行结果：
```

```
9、import numpy as np
arr10 = np.arange(0, 12).reshape(3, 4)
arr10[:2, :3].T * [1, 2]
运行结果：
```

二、代码填空（每题 3 分，共 15 分）

1、用（最好是一行）代码将下述二维列表展开成一维列表
（多于一行代码实现最高得 2 分）

```
lst = [[1, 2], [3, 4], [5, 6, 7]]
```

```
lst2 = _____
```

3、（3 分）Z-Score 标准化是特征处理的方法，对于给定的矩阵 X ，请按列对其进行标准化（ $x - \text{mean} / \text{std}$ ）

```
X = np.random.random((100, 30))
```

X_mean = _____(1 分)

X_std = _____(1 分)

X_norm = _____(1 分)

4、

```
import pandas as pd
```

```
df = pd.DataFrame( {'col1':['C1','C1','C2','C2','C2','C3','C2'],  
                    'col2':[1,2,3,3,4,6,5]})
```

要求将 col1 相同的 col2 合并到一个 list 中，

如果不记得如何生成 list 可以自定义函数;或自己命名一个函数来使用（扣 1 分）

即 df1

col1

C1 [1, 2]

C2 [3, 3, 4, 5]

C3 [6]

df1 = _____

简答题（共 30 分）

4、请简要写出离散余弦变换的基本思想及能实现有损压缩的原理（5分）

请你简要写出课程大作业里具体负责完成的工作内容和主要心得或建议。独立完成的只要写大作业题目内容，合作完成的建议写一下具体负责的工作部分。本题题面只有4分，对优秀大作业的主要完成人会用额外5分的加分来体现鼓励。

Python 程序实现（共10分）

2、Trie 实现词频统计，请补全 info 函数定义（10分）

```
class Trie:
    class Node:
        def __init__(self, char):
            """
            self.children: a map, char => Node
            self.char: char of this node
            """
            self.children = {}
            self.char = char
            self.cnt = 0

    def __init__(self):
        self.head = self.Node("")

    def find(self, key):
        node = self.head
        for char in key:
            if char in node.children:
                node = node.children[char]
            else:
                return None
        return node.cnt

    def insert(self, key):
        node = self.head
        for char in key:
```

```

        if char not in node.children:
            node.children[char] = self.Node(char)
        node = node.children[char]
        node.cnt += 1

def delete(self, key):
    node = self.head
    for char in key:
        if char in node.children:
            node = node.children[char]
        else:
            return None
    node.cnt -= 1

def info(self): # 默认无空末端节点
                # 补全代码实现遍历 Trie 树功能

```

10、深度学习部分（10）

建议出读程序对特定语句/语句块详细说明其功能和思想。也可以有程序填空。