阅资料，思考并回答下述问题：  
1解释信源编码的概念。

答：关于随机变量X的信源编码C 是从X的取值空间到的一个映射，其中表示字母表D上有限长度的字符串所构成的集合.

2信源编码的平均码长下界为熵。该结论彰显了熵可以解释为:描述复杂度--用符号串描述一个随机变量所需要的最短长度。从而到目前为止，熵有了三种解释。从这些解释中，给出数据压缩的解释。

答：Shannon借用了热力学中的名词"熵"(Entropy)来表示一条信息中真正需要编码的信息量：考虑用0和l组成的二进制数码为含有n个符号的某条信息编码。假设符号Fn在整条信息中重复出现的概率为Pn，则该符号的熵也即表示该符号所需的位数位为：En=-log(Pn)。整条信息的熵也即表示整条信息所需的位数为,E=ΣEn.用较少的位数表示较频繁出现的符号这就是数据压缩的基本准则.

3解释Kraft不等式。

答：对于D元字母表上的即时码，码字长度必定满足不等式



反之，若给定满足以上不等式的一组码字长度，则存在一个相应的即时码，其码字长度就是给定的长度。

证明：考虑每一节点均含有D个子节点的D叉树。假定数值代表码字的字符。每个码字都由树的一片叶子表示。因此，始于根节点的路径可描绘出码字中的所有字符。

码字的前缀条件表明树中无一码字是其他任一码字的祖先。因而，在这样的编码树中，每一码字都去除了他的可能成为码字的所有后代。

令为码字集中最长码字长度。考虑在树中层的所有节点，可知其中有些是码字，；有些是码字的后代；另外的既不是码字，也不是码字的后代。在树中层的码字拥有中的个后代。所有这样的后代集不相交。而且，这些集合中的总结点数必定小于等于。因此



即

.

4证明:前缀码的码长序列与满足Kraft不等式的正整数序列之间是等价的。

证明：在3中已经证明，前缀码的码长序列满足Kraft不等式。

反之，若给定任意一组满足Kraft不等式的码字长度，总可以构造出一棵编码树。将第一个深度为的节点（依字典序）标为码字1，同时除去树中属于它的所有后代。然后在剩余的节点中找出第一个深度为的节点，将其标为码字2，同时除去树中属于它的所有后代。…按此方法继续，即可构造出一个码字长度为的前缀码。

5证明二元前缀码的平均码长的下界为H(X).

证明：我们证明一般形式D元前缀码.

则

所以

令，，，

所以



即L>=H,当且仅当（即对所有的i，为整数），等号成立.

6给出Huffman编码算法与配套数据结构的形式描述.

答：Huffman码可以通过构造Huffman树来实现，构建Huffman树的算法如下：

Step 1：给定权重值集合（即概率）

Step 2：首先构造一个森林F，它有下面的n棵树构成：

1. 一个节点的二叉树，；
2. 每个只有一个权重为的根节点；

Step 3：然后，循环往复执行下列步骤，直到霍夫曼树被完全构成为止:  
 a)在森林里选取跟权重最小的两棵树   
 b)将他们合并为一棵树(通过增加一个根作为该两个节点的父结点)  
 c)合并后的树的根节点权重为两棵子树根节点权重之和  
 d)将刚才使用过的两棵树从森林中删除，将新合并的数加入到森林F里

有了霍夫曼树后，进行霍夫曼编码就是非常容易了:从霍夫曼树的根结点开始，顺着分枝往下，左边分支赋值0，右边分支赋值1(也可以反过来)，直到所有分支上都有赋值为止。此时，一个节点的编码就是从根节点到该叶子节点路径上所有赋值的连接;

7解释Huffman编码算法蕴含的算法思想。

答：

1)对所有符号按照出现和使用频率由高到低进行排序  
2)顺序取出一个符号   
3)赋予当前可用的最短编码  
4)如果还有符号没有编码，则重复第2～3步;否则结束  
霍夫曼编码符合一切贪婪的属性:每次选取使用频率最高的符号，而赋以则是当前可用的最短编码，因此，霍夫曼编码是一种贪婪策略;

8给出前缀码的两种编码思路

答：①对可数集，将随机变量值域映射为树中的叶节点集合；

②对任意构成前缀码的可数无限码字集，将随机变量值域映射为[0,1]区间的不交子区间集合；

9证明Huffman编码的最优性。

证明：设某个Huffman编码加权和为,设其中与H编码中权重Pi的深度Ai, Pj的深度Aj，不妨设Pi>Pj，则Ai<Aj.

假设调换这两个节点，可得到更小的期望长度，则有Pi\*Aj+Pj\*Ai<Pi\*Ai+Pj\*Aj，整理得Pj(Aj-Ai)>Pi(Aj-Ai)，即Pj>Pi, 这与Pi>Pj矛盾.故假设错误，因此不存在更优的编码。

10复习可导函数的有约束优化中的Lagrangian乘子法。  
（答案提交方式：电子稿，电子邮箱,infosecynu2015@163.com)