第 5 讲 S

组合数学

计数的基本原则

1. 相等规则

如果 A 和 B 是有限集合，且存在从 A 到 B 的双射，则 $|A| = |B|$

示例

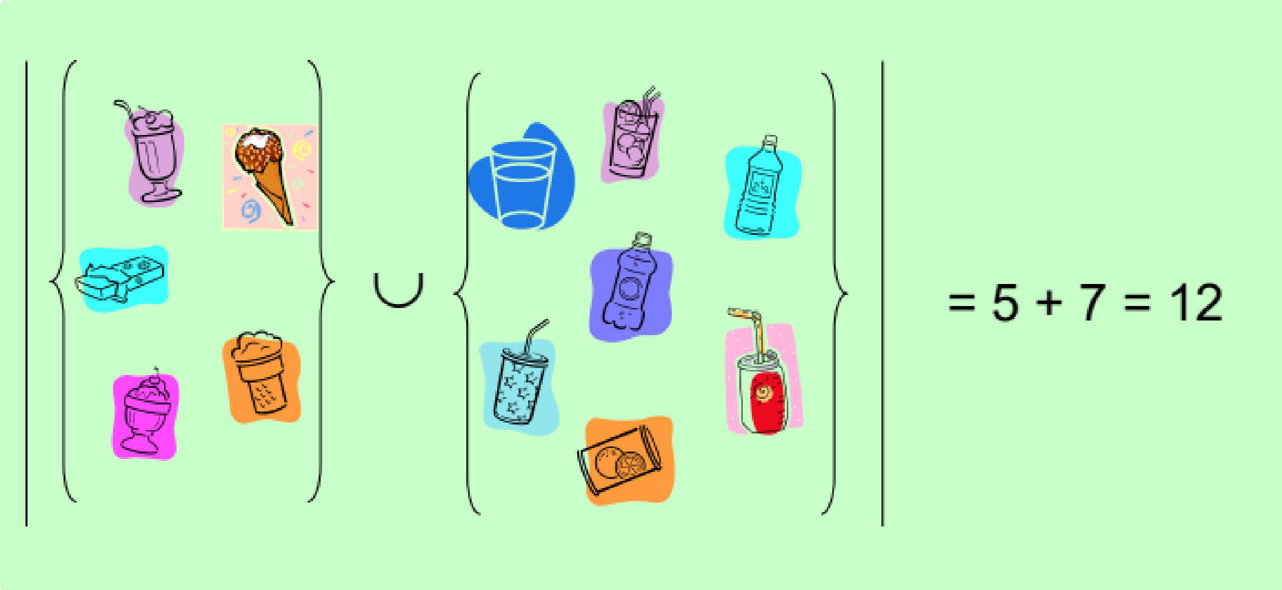
有 5 种冰淇淋

和 7 种饮料

选择其中一样东西有多少种方法?

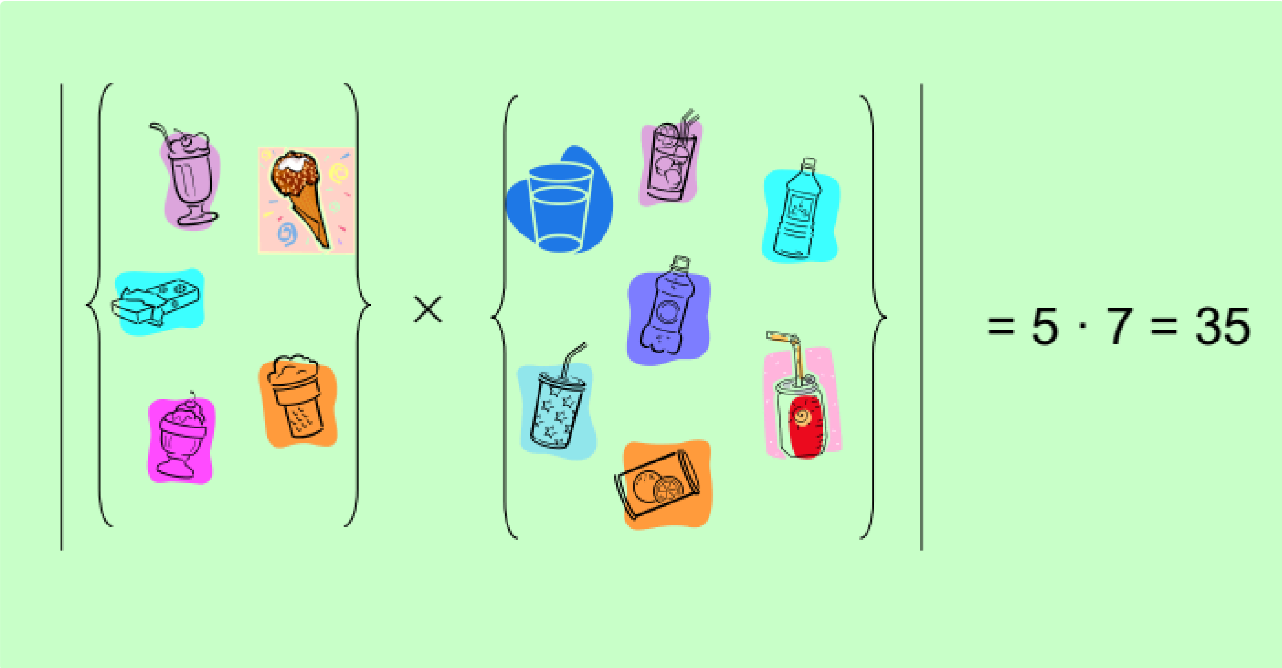
加法规则

如果 A 和 B 是不相交的有限集合（即 $A \cap B = \emptyset$），则 $|A \cup B| = |A| + |B|$。



乘法规则

如果 A 和 B 是有限集合，则 $|A \times B| = |A| \cdot |B|$ 。



为了证明乘法规则，我们考虑两个集合

$A = \{a_1, a_2, ..., a_n\}$ ，$B = \{b_1, b_2, ..., b_k\}$ 。

我们构建一个决策树来选择 $(x, y) \in A \times B$ 。

根节点包含问题"x 等于什么？"

从根节点出发有 n 个箭头，对应可能的答案 $a_1, a_2, ..., a_n$ 。这些箭头指向第一层节点。

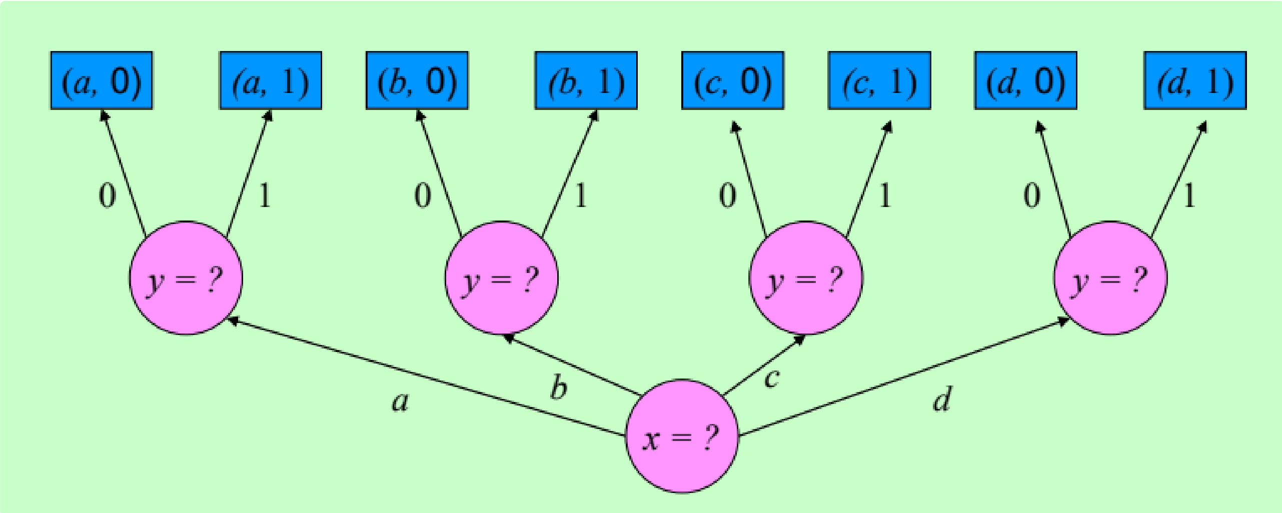
每个第一层节点包含问题："y 等于什么？"从这些节点出发有 k 个箭头，对应可能的答案 $b_1, b_2, ..., b_k$ 。这些箭头指向第二层节点，即叶子节点。

每个叶子节点对应 $A \times B$ 中的一个对。

有 n 个第一层节点，每个节点有 k 个出射箭头。因此，树中有 nk 个叶子节点。

叶子节点和 $A \times B$ 中的对之间存在一一对应关系。因此，对的数量等于 $nk = |A| \cdot |B|$ 。

示例： $A = \{a, b, c, d\}$ ， $B = \{0, 1\}$



这两个规则可以推广到任意数量的集合。

如果 $A_1, A_2, \ldots, A_k$ 是互不相交的有限集合，则

$|A_1 \cup A_2 \cup \ldots \cup A_k| = |A_1| + |A_2| + \ldots + |A_k|$ 。

如果 $A_1, A_2, \ldots, A_k$ 是有限集合，则

$|A_1 \times A_2 \times \ldots \times A_k| = |A_1| \cdot |A_2| \cdot \ldots \cdot |A_k|$ 。

作为乘法规则的一个特例，当 $A_1 = A_2 = \ldots = A_k$ 时，我们得到：

定理：由集合 A（基数为 k）的元素组成的长度为 n 的序列的数量等于 k^n **。**

顺序选择原理

设序列 $x_1, x_2, \ldots, x_k$ 是通过依次选择元素 $x_1, x_2, \ldots, x_k$ 形成的，其中：

- 元素 x_1 可以通过 n_1 种方式选择；
- 对于任意 x_1 ，元素 x_2 可以通过 n_2 种方式选择；
- 对于任意 x_1, x_2 ，元素 x_3 可以通过 n_3 种方式选择；

...

- 对于任意 $x_1, x_2, \ldots, x_{k-1}$ ，元素 x_k 可以通过 n_k 种方式选择。

那么，整个序列可以通过 $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ 种方式选择。

这个原理可以通过决策树来证明。

词语

字母表是一组字母（符号）的集合。

通过按某种顺序写出字母，我们得到一个**词**。

例如，从字母表 $A = \{a, b, c\}$ 中，我们可以形成词：acb, aaaa, bab, cca, b 等。

存在一个**空词**，不包含任何字母。它用 λ 表示。

实际上，词是一个不带括号和逗号的符号序列。

有时，"词"这个术语也被称为"**字符串**"。

词的**长度**是指它包含的字母数量。

字母表 A 中所有长度为 n 的词的集合记为 A^n 。

特别地， $A^0 = \{\lambda\}$ ， $A^1 = A$ 。

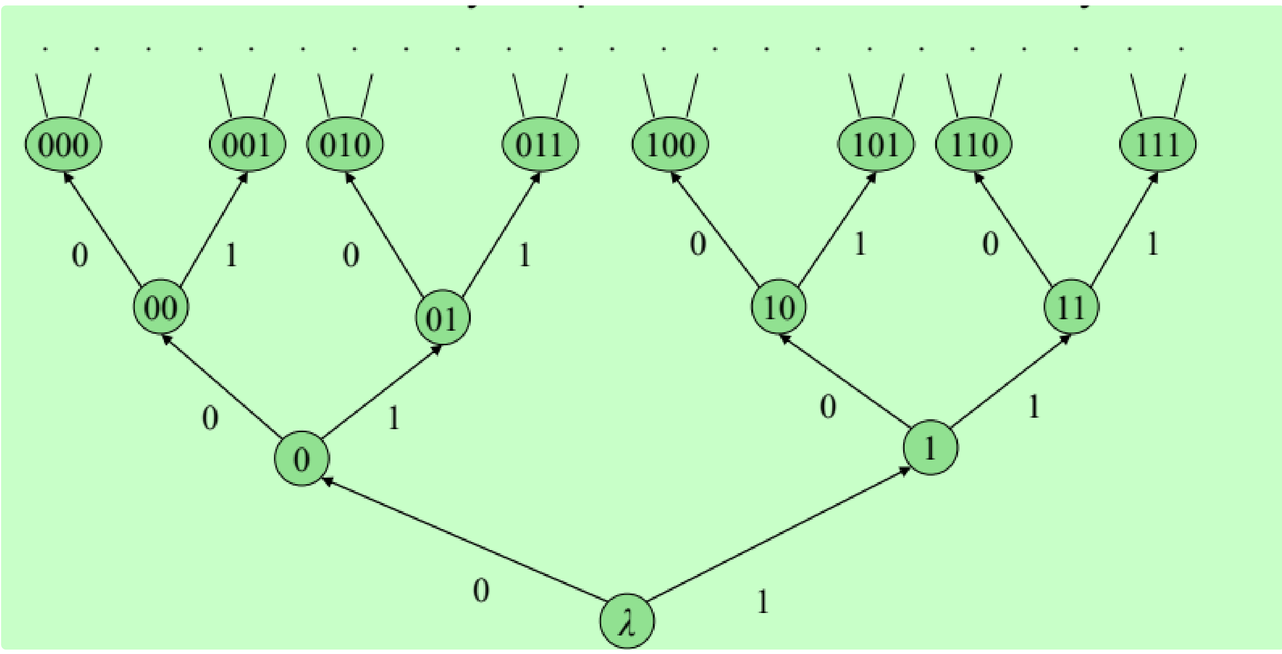
在基数为 k 的字母表中，长度为 n 的词的数量等于相应序列的数量，即 k^n 。

字母表 A 中所有词的集合记为 A^* 。

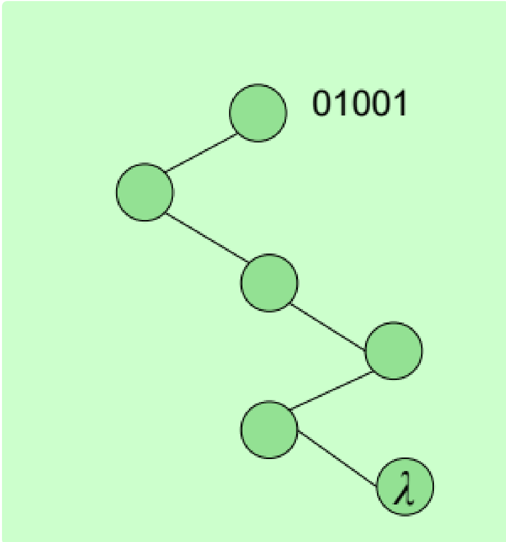
因此， $A^* = A^0 \cup A^1 \cup A^2 \cup \ldots \cup A^n \cup \ldots$

二进制词的表示

字母表 $\{0, 1\}$ 中的词（二进制词）可以用二叉树表示。字母 0 和 1 与从节点出发的两个箭头相对应。

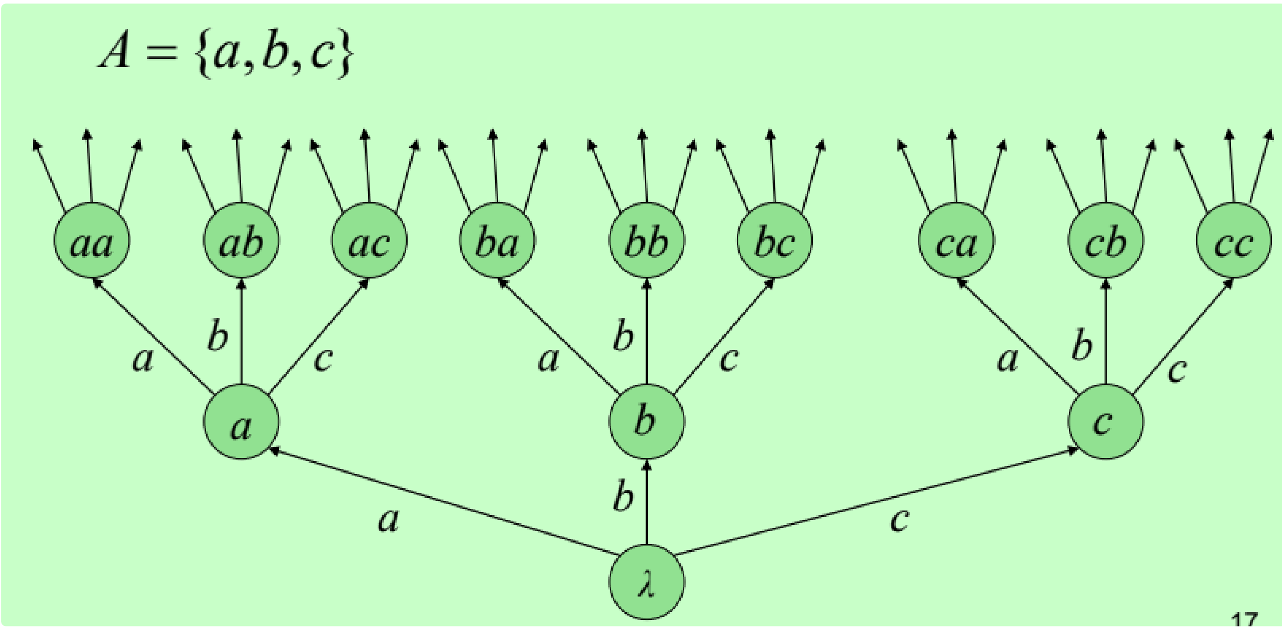


左箭头始终为 0，右箭头始终为 1。沿着从根到某个节点的路径移动，可以读出一个词。这个词由该节点表示。



所有二进制词的集合由一个无限二叉树表示。对于每个词，树中都有一个唯一的节点表示该词。

利用树可以表示任意字母表中的单词。如果字母表由 k 个字母组成，则使用 k 叉树，其中每个节点有 k 个分支，分别用不同的字母标记。



字典序

为什么在世界国家列表中，澳大利亚（Australia）排在奥地利（Austria）之前?

这是因为列表是按字母顺序排列的。这意味着什么呢?

设 A 是一个有限字母表，其上定义了线性顺序（字母表上的字母顺序）。

我们用 $\leq$ 表示这个顺序。

如果 $x \leq y$ 且 $x \neq y$，则写作 $x < y$。

字典序是将字母顺序扩展到集合 A^* 上。

对于字典序，我们使用相同的符号 $\leq$，$<$ 。

首先，我们为相同长度的单词定义字典序。

设 $\alpha = a_1 a_2 \ldots a_n$ 和 $\beta = b_1 b_2 \ldots b_n$ 是 A^n 中的**单词**。

单词 α 在字典序中小于单词 β，记作

$$\alpha < \beta \text{ 。}$$

如果存在某个 i：

$$a_1 = b_1, a_2 = b_2, \ldots, a_{i-1} = b_{i-1},$$

$$a_i < b_i$$

传递性证明：$\alpha \leq \beta, \beta \leq \gamma \Rightarrow \alpha \leq \gamma$

如果 $\alpha = \beta$ 或 $\beta = \gamma$，结论显然成立。

假设 $\alpha \neq \beta$ 且 $\beta \neq \gamma$。

设 $\alpha = a_1 a_2 \ldots a_n$，$\beta = b_1 b_2 \ldots b_n$，$\gamma = c_1 c_2 \ldots c_n$ 。

那么：

$\alpha \leq \beta \Rightarrow$ 存在 i，使得 $a_1 a_2 \ldots a_{i-1} = b_1 b_2 \ldots b_{i-1}$ 且 $a_i < b_i$，

$\beta \leq \gamma \Rightarrow$ 存在 k，使得 $b_1 b_2 \ldots b_{k-1} = c_1 c_2 \ldots c_{k-1}$ 且 $b_k < c_k$ 。

考虑两种情况：$i \leq k$ 和 $i > k$ 。

- 当 $i \leq k$ 时：

  > $a_1 = b_1 = c_1$，
  >
  > $a_2 = b_2 = c_2$，
  >
  > ...,
  >
  > $a_{i-1} = b_{i-1} = c_{i-1}$，
  >
  > $a_i < b_i \leq c_i$
  >
  > 因此，$a_1 = c_1$，...，$a_{i-1} = c_{i-1}$，$a_i < c_i$，
  >
  > 所以 $\alpha < \gamma$。

- 当 $i > k$ 时，类似地：

  > $a_1 = b_1 = c_1$，
  >
  > $a_2 = b_2 = c_2$，
  >
  > ...,
  >
  > $a_{k-1} = b_{k-1} = c_{k-1}$，
  >
  > $a_k = b_k < c_k$

再次得到 $\alpha < \gamma$ 。

字典序是一个线性顺序。

让我们按字典序列出长度为 3 的二进制单词：

单词	编号
000	0
001	1
010	2
011	3
100	4
101	5
110	6
111	7

$\text{编号}(a_1a_2a_3) = 4a_1 + 2a_2 + a_3 = a_1 \cdot 2^2 + a_2 \cdot 2^1 + a_3 \cdot 2^0$

一般情况：长度为 n 的二进制单词

与单词 $\alpha = a_1a_2...a_n$ 关联的整数为：

$$N(\alpha) = a_1 \cdot 2^{n-1} + a_2 \cdot 2^{n-2} + ... + a_{n-1} \cdot 2^1 + a_n \cdot 2^0$$
$$= \sum_{i=1}^{n} 2^{n-i} \cdot a_i$$

最小值为

$N(000...0) = 0$,

最大值为

$N(111...1) = 2^{n-1} + 2^{n-2} + ... + 2^1 + 2^0 = 2^n - 1$.

定理：函数 N 是从 $\{0,1\}^n$ 到 $\{0,1,...,2^n-1\}$ 的双射。

证明：首先证明单射性。

更准确地说，我们证明如果 $\alpha < \beta$ ，则 $N(\alpha) < N(\beta)$ 。

确实，设 $\alpha = a_1a_2...a_n$ ，$\beta = b_1b_2...b_n$ 且 $\alpha < \beta$ 。

那么存在 k ，使得 $a_1a_2...a_{k-1} = b_1b_2...b_{k-1}$ ，

$a_k < b_k$ ，即 $a_k = 0$ ，$b_k = 1$ 。

那么

$$N(\alpha) - N(\beta) = (a_1 \cdot 2^{n-1} + ... + a_n \cdot 2^0) - (b_1 \cdot 2^{n-1} + ... + b_n \cdot 2^0)$$

$$= (a_{k+1} \cdot 2^{n-k-1} + ... + a_n \cdot 2^0) - (2^{n-k} + b_{k+1} \cdot 2^{n-k-1} + ... + b_n \cdot 2^0)$$

$$\leq (2^{n-k-1} + ... + 2^0) - 2^{n-k} = 2^{n-k} - 1 - 2^{n-k}) = 1$$

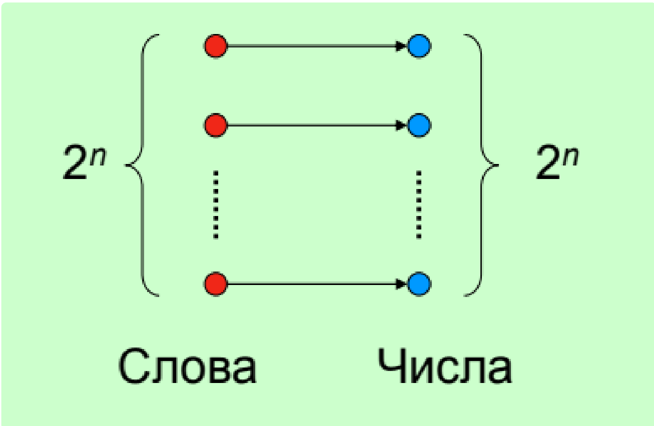
因此,

$$N(\alpha) < N(\beta)$$

所以，函数 N 是单射的。

集合 $\{0, 1\}^n$ 和 $\{0, 1, ..., 2^n - 1\}$ 有相同的基数 2^n 。

由此可以推断，这个函数是满射的。



计算 N^{-1} 函数

如何根据单词的编号计算单词本身?

需要找到 $a_1, a_2, ..., a_n$ ，如果已知 x 。

假设单词的编号为

$$x = N(\alpha) = a_1 \cdot 2^{n-1} + a_2 \cdot 2^{n-2} + ... + a_{n-1} \cdot 2^1 + a_n \cdot 2^0$$

需要找到 $a_1, a_2, ..., a_n$ ，如果已知 x 。

实际上，这涉及计算数字 x 的二进制表示。

我们从第一个字母 a_1 开始。

如果 $a_1 = 0$ ，那么 $N(\alpha) \leq 2^{n-2} + \cdots + 2^0 = 2^{n-1} - 1$,

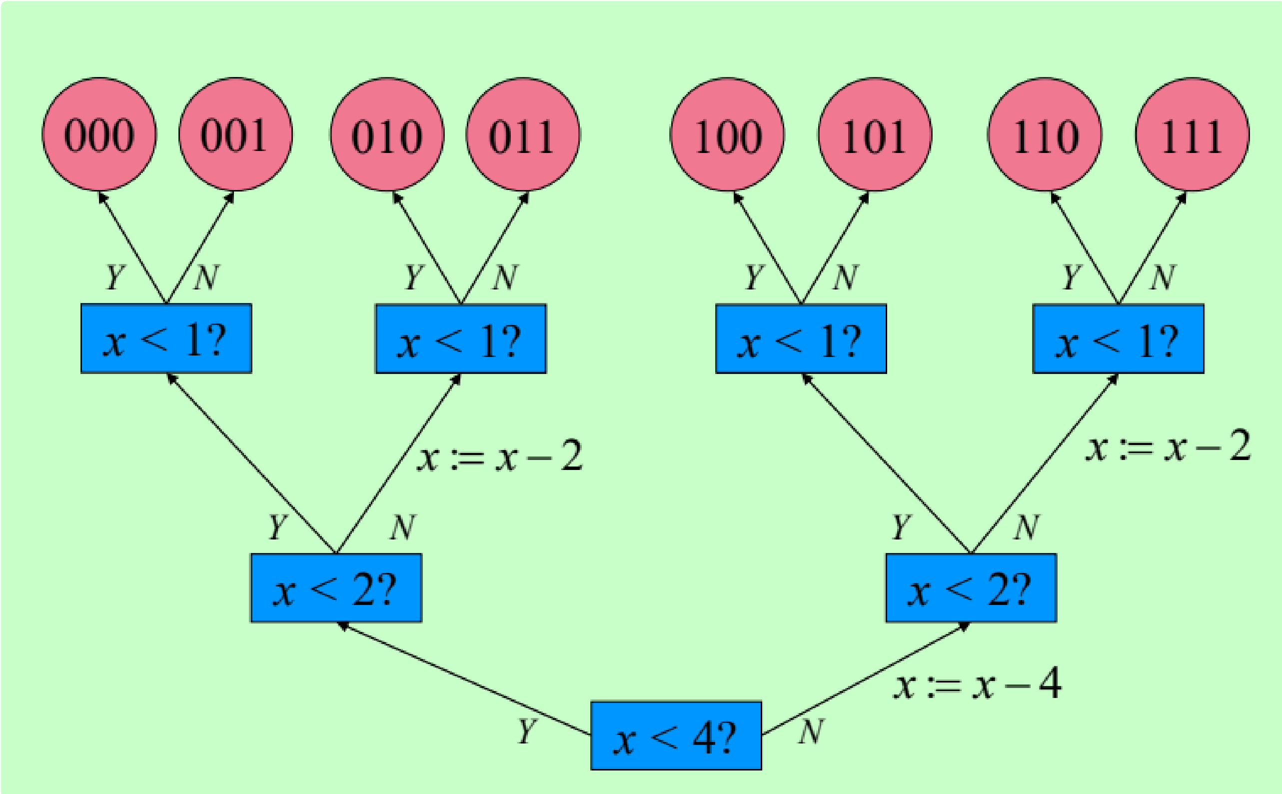
如果 $a_1 = 1$,

因此，我们通过比较 x 和 2^{n-1} 来确定 a_1 :

如果 $x \geq 2^{n-1}$ ，则 $a_1 = 1$ ，否则 $a_1 = 0$ 。

然后，我们从 x 中减去 $a_1 \cdot 2^{n-1}$ ，并以类似的方式通过与 2^{n-2} 比较来确定 a_2 ，以此类推。

对于 $n = 3$ ，这个算法可以用决策树表示:



在一般情况下，可以用伪代码描述如下：

输入： $n \in \mathbb{N}$, $x \in \mathbb{N}_0$, $x < 2^n$

输出： $a_1, a_2, \ldots, a_n$, $a_i \in \{0,1\}$ 对于 $i = 1, 2, \ldots, n$

for $i = 1$ **to** n **do**
  if $x < 2^{n-i}$
  then $a_i := 0$
  else $\{ a_i := 1;\quad x := x - 2^{n-i} \}$

例如： $x = 11$, $n = 4$

i	问题	a_i	x
1	11 < 8 ?	1	3
2	3 < 4 ?	0	3
3	3 < 2 ?	1	1
4	1 < 1 ?	1	0

字典序的一般定义

对于任意（可能不同）长度的单词，字典序定义如下：

单词 $\alpha = a_1 a_2 \ldots a_n$ 在**字典序**中小于单词 $\beta = b_1 b_2 \ldots b_m$ ，如果：

存在 i ，使得 $a_1 = b_1$, ..., $a_{i-1} = b_{i-1}$, $a_i < b_i$

或者 $n < m$ 且 $a_1 = b_1$, ..., $a_n = b_n$ 。

例如： $A = \{a, b, c\}$ ， $a < b < c$ 。那么：

bcabc < bcccaab,

acca < accabca.