第3次作业

9.1-1：证明：在最坏情况下，找到n 个元素中第二小元素需要 n+[**lg**n]−2 次比较。（提示：以同时找最小元素。）

以锦标赛方式比较元素——将其两个一组进行比较，然后以同样的方式对**获胜者**进行比较。需要跟踪潜在的赢家所参与的每次“赛事”。

通过n−1 次比较确定最终赢家。而第二小元素就在比赛输于最小元素的[**lg**n] 中——其中每个元素都是在所参与的最后一次赛事中失利。因此要找到最小元素还须[**lg**n]−1次比较。

9.3-1 在算法 SELECT 中，输入元素被分为每组 5 个元素。如果它们被分为每组 7 个元素，该算法仍然会是线性时间吗？证明：如果分成每组 3 个元素，SELECT 的运行时间不是线性的。

1）每组7个元素是可以保证线性时间的。

可以证明，当r=7时，小于mm或大于mm的元素数至少是2n/7-8个： $4\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{7}\right\rceil \right\rceil - 2\right) \geq \frac{2n}{7} - 8$，

于是有： $T(n) \leq T(\lceil n/7 \rceil) + T(5n/7 + 8) + O(n)$

可以证明T(n)=O(n)

2) 每组3个元素不能保证线性时间。

可以证明，当r=3时，小于mm或大于mm的元素数至少是2n/3-4个： $2\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{3}\right\rceil \right\rceil - 2\right) \geq \frac{n}{3} - 4$

则有：$T(n) \leq T(\lceil n/3 \rceil) + T(2n/3 + 4) + O(n)$

而n/3+2n/3+4>n非线性解。

9.3-5：假设你已经有了一个最坏情况下是线性时间的用于求解中位数的“黑箱”子过程。设计一个能在线性时间内解决任意顺序统计量的选择问题算法。

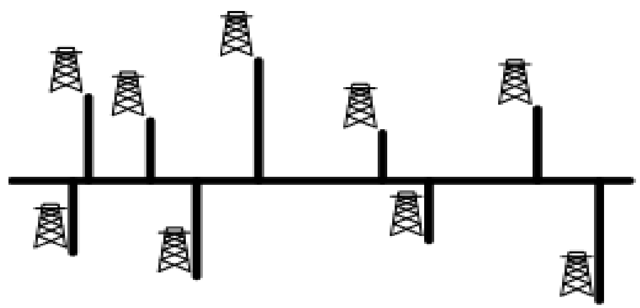
```
SELECT'(A, p, r, i)
  if p == r
      return A[p]
  x = MEDIAN(A, p, r)
  q = PARTITION(x)
  k = q - p + 1
  if i == k
      return A[q]
  elseif i < k
      return SELECT'(A, p, q - 1, i)
  else return SELECT'(A, q + 1, r, i - k)
```

如果i = ⌈n/2⌉，则只需调用一次子过程，显然是线性时间。

否则，只需针对所划分成的两部分中的一个调用子过程（视i的大小而定），因此有如下递归式：$T(n) = T(n/2)+O(n)$

由主定理可知上限为$O(n)$。

9.3-9：Olay教授是一家石油公司的顾问。这家公司正在计划建造一条从东到西的大型输油管道，这一管道将穿越一个有n 口油井的油田。公司希望每口油井都有一条管道支线沿着最短路径连接到主管道（方向或南或北），如下图所示。给定每口油井的x 和y 坐标，教授应该如何选择主管道的最优位置，使得各支线的总长度最小？证明：该最优位置可以在线性时间内确定。



如果n 是奇数，则选取所有油井y 坐标的中位数,作爲主管道的y坐标，即主管道穿过此油井。这样主管道两侧的油井数目相同。对于任两口油井而言，只要主管道在他们中间通过，那么这两口油井的支线管道总长度是不变的。

如果n是偶数，则需要所有油井y坐标的两个中位数，主管道的y坐标在这两个y坐标中间即可

选择中位数的原因可以证明。由于求解中位数的问题可以在线性时间确定

9-2 （带权中位数） 对分别具有正权重 w_1，w_2，…，w_n，且满足 $\sum_{i=1}^{n} w_i = 1$ 的 n 个互异元素 x_1，x_2，…，x_n 来说，带权中位数 x_k(较小中位数)是满足如下条件的元素：

$$\sum_{x_i<x_k} w_i < \frac{1}{2}$$

和

$$\sum_{x_i>x_k} w_i \leqslant \frac{1}{2}$$

例如，如果元素是 0.1，0.35，0.05，0.1，0.15，0.05，0.2，并且每个元素的权重等于本身(即对所有 $i=1$，2，…，7，都有 $w_i=x_i$)，那么中位数是 0.1，而带权中位数是 0.2。

a. 证明：如果对所有 $i=1$，2，…，n 都有 $w_i=1/n$，那么 x_1，x_2，…，x_n 的中位数就是 x_i 的带权中位数。

b. 利用排序，设计一个最坏情况下 $O(n\lg n)$时间的算法，可以得到 n 个元素的带权中位数。

c. 说明如何利用像 9.3 节的 SELECT 这样的线性时间中位数算法，在 $\Theta(n)$最坏情况时间内求出带权中位数。

a) 根据带权中位数的定义即可得，当每个数的权重都为1/n时，带权中位数就是排序后位于中间的那个数，也即中位数就是带权中位数。

b）先排序，花O(nlogn)的时间，然后顺序搜索：从排序后的第一元素开始，累加权值sum-w，直到某个元素x_i，$\text{sum-}w_{1\sim i-1}<1/2$，$\text{sum-}w_{1\sim i-1}+w_i \geqslant 1/2$，则$x_i$就是该序列的带权中位数.

c）以下程序给出在Θ(n)时间内找带权中位数的算法

WEIGHTED-MEDIAN(X)

if $n == 1$
  return x_1
elseif $n == 2$
  if $w_1 \geq w_2$
    return x_1
  else return x_2
else find the median x_k of $X = \{x_1, x_2, \ldots, x_n\}$
  partition the set X around x_k
  compute $W_L = \sum_{x_i < x_k} w_i$ and $W_G = \sum_{x_i > x_k} w_i$
  if $W_L < 1/2$ and $W_G < 1/2$
    return x_k
  elseif $W_L > 1/2$
    $w_k = w_k + W_G$
    $X' = \{x_i \in X : x_i \leq x_k\}$
    return WEIGHTED-MEDIAN(X')
  else $w_k = w_k + W_L$
    $X' = \{x_i \in X : x_i \geq x_k\}$
    return WEIGHTED-MEDIAN(X')

如果只有两个元素，直接二者的权值比较即可。以权值较大的一个作为带权中位数。

按常规的方法找中位数（普通中位数，不是带权的中位数）。

以x_k为界分为左、右两个子集（左子集的所有元素不大于x_k，右子集的所有元素不小于x_k），然后计算左、右子集的权值之和：W_L和W_G，如果W_L和W_G都小于1/2，则x_k就是要求的带权中位数。

左侧

右侧

否则，则在局部权值和较大的子集：左子集L或右子集R里重复上述过程，但注意，x_k参与下一步在L或R中的搜索，并且在递归进行下一次搜索之前将另一个子集R或L的权值之和W_R或W_L累加到w_k上，这样k元素的新权值就代表了将舍去的另一半子集中所有元素的权值之和，后续计算就可以计算出“全局”权值的和，而不会丢失被舍去的子集的元素的权值。

- 可以证明上述算法的时间是：

$$T(n) = T(n/2+1) + \Theta(n).$$

- 所以总的时间是T(n)= Θ (n).

邮局位置问题的定义如下：给定权重分别为 w_1，w_2，…，w_n 的 n 个点 p_1，p_2，…，p_n，我们希望找到一个点 p（不一定是输入点中的一个），使得 $\sum_{i=1}^{n} w_i d(p, p_i)$ 最小，这里 $d(a, b)$ 表示点 a 与 b 之间的距离。

d. 证明：对一维邮局位置问题，带权中位数是最好的解决方法，其中，每个点都是一个实数，点 a 与 b 之间的距离是 $d(a, b) = |a-b|$。

d) The property that $\sum_{x_i > x} w_i \leq 1/2$ implies that $\sum_{x \leq x_i} w_i > 1/2$. This fact, combined with $x - y > 0$ and $\sum_{x > x_i} w_i < 1/2$, yields that $f(y) - f(x) > 0$.

When $x > y$, we again bound the quantity $|y - x_i| - |x - x_i|$ from below by examining three cases:

1. $x_i \leq y < x$: Here, $|y - x_i| + |x - y| = |x - x_i|$ and $|x - y| = x - y$, which imply that $|y - x_i| - |x - x_i| = -|x - y| = y - x$.
2. $y \leq x_i < x$: Here, $|y - x_i| \geq 0$ and $|x - x_i| \leq x - y$, which imply that $|y - x_i| - |x - x_i| \geq -(x - y) = y - x$.
3. $y < x \leq x_i$. Here, $|x - y| + |x - x_i| = |y - x_i|$ and $|x - y| = x - y$, which imply that $|y - x_i| - |x - x_i| = |x - y| = x - y$.

Separating out the first two cases, in which $x > x_i$, from the third case, in which $x \leq x_i$, we get

$$
\begin{aligned}
f(y) - f(x) &= \sum_{i=1}^{n} w_i (|y - x_i| - |x - x_i|) \\
&\geq \sum_{x > x_i} w_i (y - x) + \sum_{x \leq x_i} w_i (x - y) \\
&= (x - y) \left(\sum_{x \leq x_i} w_i - \sum_{x > x_i} w_i \right) .
\end{aligned}
$$

The property that $\sum_{x_i > x} w_i \leq 1/2$ implies that $\sum_{x \leq x_i} w_i > 1/2$. This fact, combined with $x - y > 0$ and $\sum_{x > x_i} w_i < 1/2$, yields that $f(y) - f(x) > 0$.

当x<y时，可类似的进行分类讨论。

邮局位置问题的定义如下：给定权重分别为 w_1，w_2，…，w_n 的 n 个点 p_1，p_2，…，p_n，我们希望找到一个点 p(不一定是输入点中的一个)，使得 $\sum_{i=1}^{n} w_i d(p,p_i)$ 最小，这里 $d(a,b)$ 表示点 a 与 b 之间的距离。

e. 请给出二维邮局位置问题的最好解决方法：其中的点是(x,y)的二维坐标形式，点$a=(x_1,y_1)$与$b=(x_2,y_2)$之间的距离是 Manhattan 距离，即 $d(a,b)=|x_1-x_2|+|y_1-y_2|$。

$$f(x,y)=\sum_{i=1}^{n} w_i\left(|x-x_i|+|y-y_i|\right)$$

$$\begin{aligned}
\min_{x,y} f(x,y) &= \min_{x,y}(g(x)+h(y)) \\
&= \min_{x}\left(\min_{y}(g(x)+h(y))\right) \\
&= \min_{x}\left(g(x)+\min_{y} h(y)\right) \\
&= \min_{x} g(x)+\min_{y} h(y)\,.
\end{aligned}$$

故分别选取x坐标和y坐标的带权中位数即为最好的解法。

- 思考题

分金币（Spreading the Wealth, UVa 11300）

圆桌旁坐着n个人，每人有一定数量的金币，金币总数能被n整除。每个人可以给他左右相邻的人一些金币，最终使得每个人的金币数目相等。你的任务是求出被转手的金币数量的最小值。比如，n=4，且4个人的金币数量分别为1,2,5,4时，只需转移4枚金币（第3个人给第2个人两枚金币，第2个人和第4个人分别给第1个人1枚金币）即可实现每人手中的金币数目相等。

解：设n个人分别拥有的金币数为$a_1, a_2, \ldots, a_n$

最终每个人所拥有金币数为A,

令x_i为i给i+1的金币数(i=1,2,...,n-1) (x_i为负时即为i+1给i金币)，x_n代表n给1的金币数

则有$A = \sum_{i=1}^{n} a_i / n$

故有

$$
\begin{aligned}
a_1 + x_n - x_1 &= A \\
a_2 + x_1 - x_2 &= A \\
a_3 + x_2 - x_3 &= A \\
&\cdots \cdots \\
a_n + x_{n-1} - x_n &= A
\end{aligned}
$$

转手出的金币数量总数为

$$x_2 = x_1 - (A - a_2)$$

$$x_3 = x_1 - (2 * A - a_2 - a_3)$$

$$\dots\dots$$

结果就是求$|x_1|+|x_1-b_1|+|x_1-b_2|+\dots+|x_1-b_{n-1}|$的最小值，取中位数即可。