6.006
Spring 2008
L10

Sorting III

Sorting lower bounds
- Decision trees

Linear-time sorting
- Counting sort

Readings : CLRS 8.1, 8.2, 8.3, 8.4

Comparison Sorting

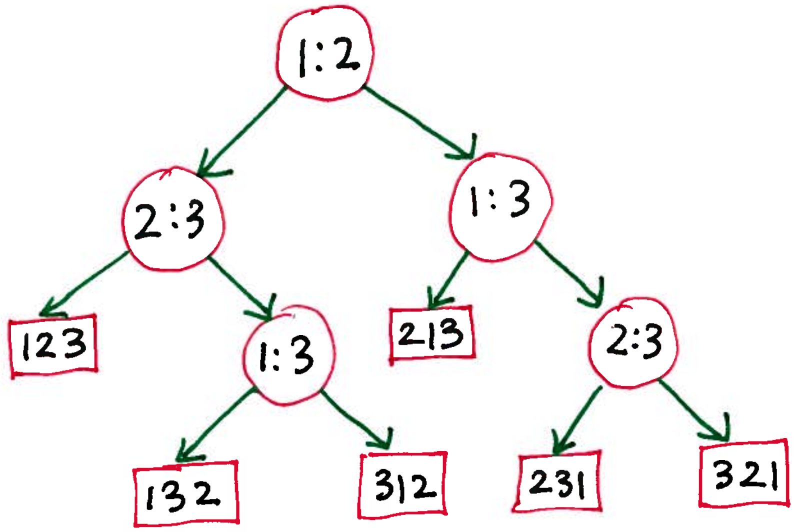
Insertion sort, merge sort and heap sort are all comparison sorts

Best worst-case running time we know is $O(n \lg n)$

Can we do better?

Decision-Tree Example

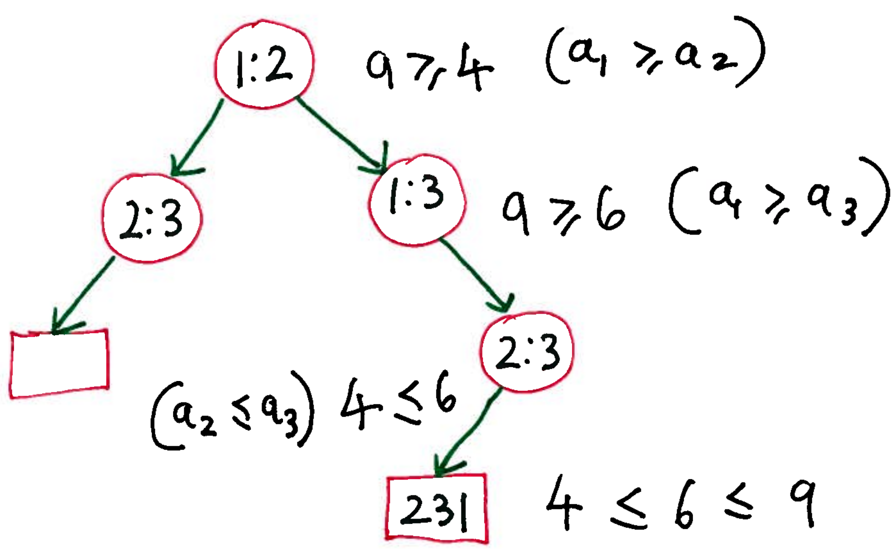
Sort $\langle a_1, a_2, \ldots a_n \rangle$



Each internal node labeled $i:j$, compare a_i and a_j, go left if $a_i \le a_j$, go right otherwise

EXAMPLE

Sort $\langle a_1, a_2, a_3 \rangle = \langle 9, 4, 6 \rangle$



each leaf contains a permutation, i.e., a total ordering

DECISION TREE MODEL

(can model execution of any comparison sort)

One tree size for each input size n

Running time of algo: length of path taken

Worst-case running time: height of the tree

In order to sort, we need to generate a total ordering of the elements!

Theorem

Any decision tree that can sort n elements must have height $\Omega(n \lg n)$.

Proof: Tree must contain $\geq n!$ leaves since there are $n!$ possible permutations.

A height-h binary tree has $\leq 2^h$ leaves. Thus, $n! \leq 2^h$

$$\Rightarrow h \geq \lg(n!) \quad \left(\geq \lg\left(\left(\frac{n}{e}\right)^n\right) \text{ Stirling}\right)$$

$$\geq n\lg n - n\lg e$$

$$= \Omega(n \lg n)$$

$\boxtimes$

Sorting in Linear Time

Counting Sort : no comparisons between elements

Input: $A[1...n]$ where $A[j] \in \{1, 2, .. k\}$
Output: $B[1..n]$ sorted
Auxiliary storage : $C[1..k]$

INTUITION

Since elements are in the range $\{1, 2 .. k\}$ imagine collecting all the j's such that $A[j] = 1$, then the j's such that $A[j] = 2$, etc.

Don't compare <u>elements</u>, so it is not a comparison sort!

$A[j]$'s <u>index</u> into appropriate positions

PSEUDO CODE & ANALYSIS

$\Theta(k)$ { for $i \leftarrow 1$ to k
do $C[i] = 0$

$\Theta(n)$ { for $j \leftarrow 1$ to n
do $C[A[j]] = C[A[j]] + 1$

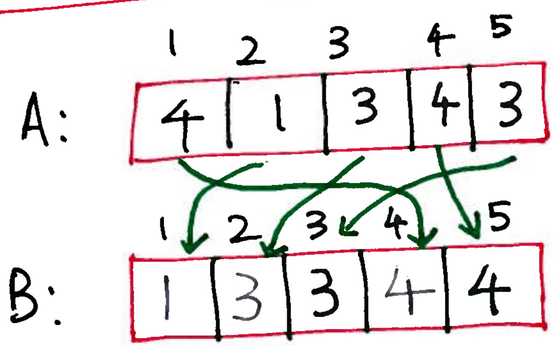
$\Theta(k)$ { for $i \leftarrow 2$ to k
do $C[i] = C[i] + C[i-1]$

$\Theta(n)$ { for $j \leftarrow n$ downto 1
do $B[C[A[j]]] = A[j]$
$C[A[j]] = C[A[j]] - 1$

$\Theta(n+k)$

Note: records may be associated with the $A[i]$'s

EXAMPLE



C:

1	2	3	4
0	0	0	0

C:

1	0	2	2

C:

1	2	3	4
1	1	$\cancel{3}$ 2	$\cancel{5}$ 4

$A[n] = A[5] = 3$
$C[3] = 3$
$B[3] = A[5] = 3$, $C[3]$ decr.

$A[4] = 4$
$C[4] = 5$
$B[5] = A[4] = 4$, $C[4]$ decr.

and so on...