Decision Tree
for Comparision Sorting

Instructor: Shizhe Zhou

Course Code:00125401

Comparison Sorting

- Insertion Sort
- Merge Sort
- QuickSort
- Heap Sort
  - The best worst-case running time that we've seen for comparison is $O(n\lg n)$.
  - Can it be further improved?

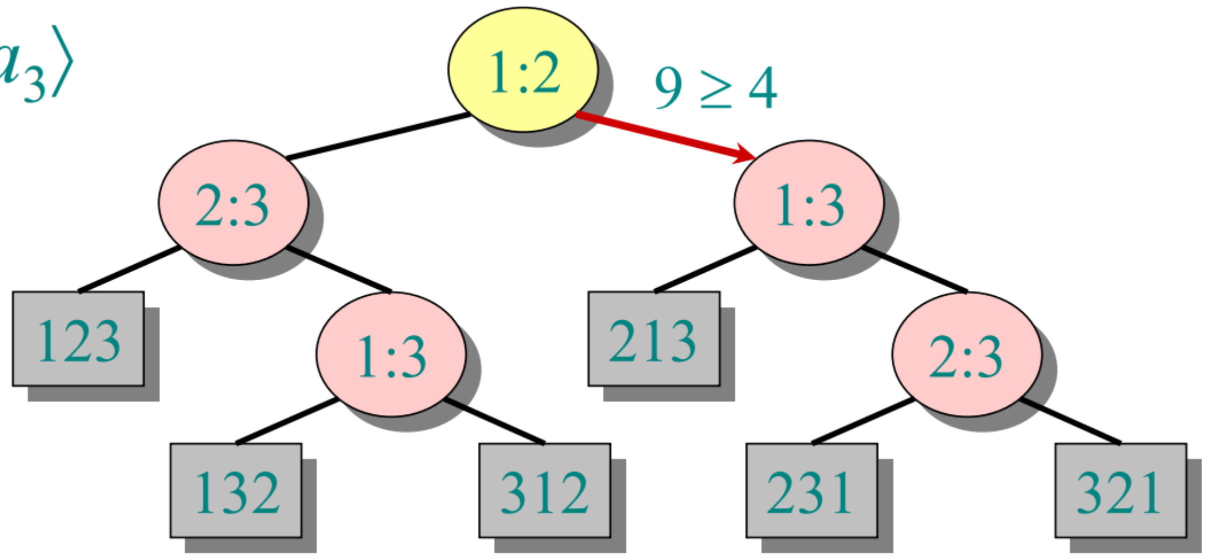
2叉分流决策

3叉分流决策

决策树可以是二叉的，也可以是多叉的。

Decision Tree Example

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

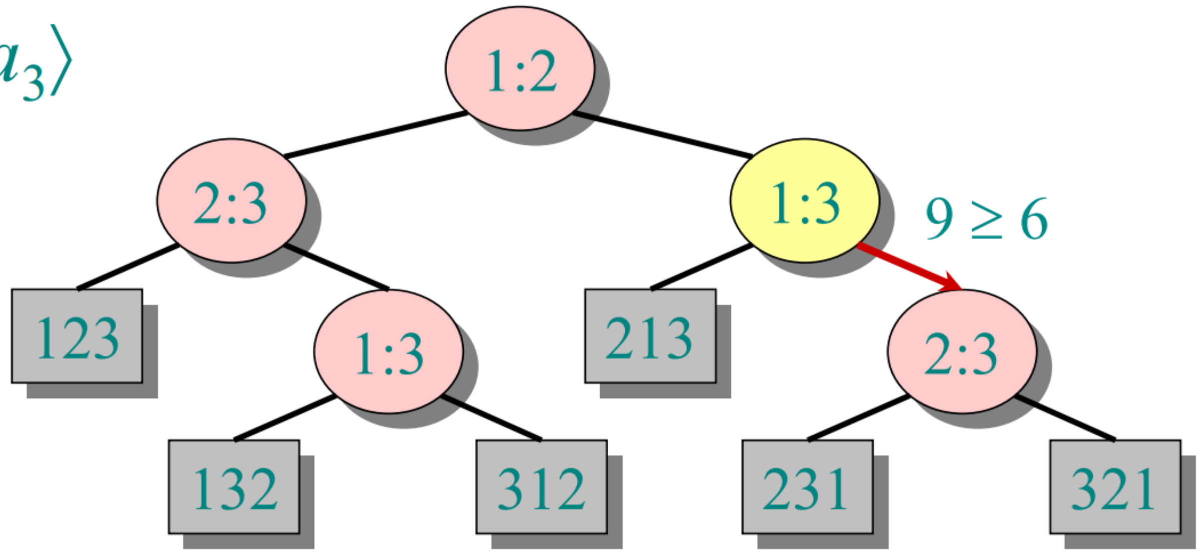


Each internal node is labeled $i{:}j$ for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i \leq a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision Tree Example

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

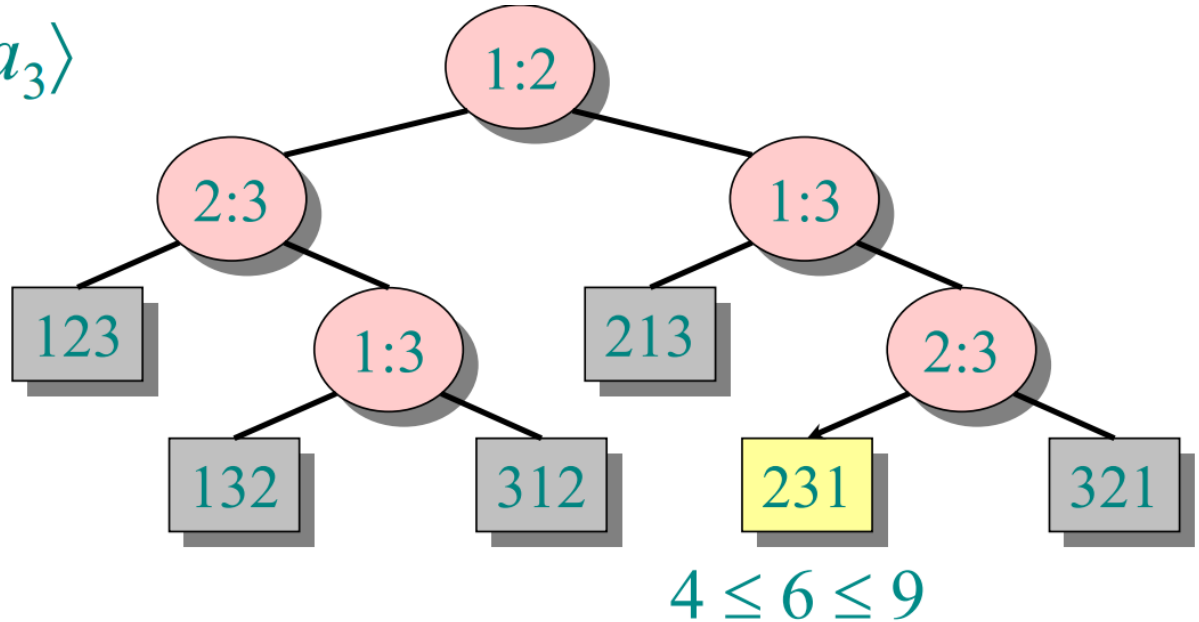


Each internal node is labeled $i{:}j$ for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i \leq a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision Tree Example

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



Each leaf contains a permutation $\langle \pi(1), \pi(2), \ldots, \pi(n) \rangle$ to indicate that the ordering $a_{\pi(1)} \leq a_{\pi(2)} \leq \cdots \leq a_{\pi(n)}$ has been established.

Decision Tree Analysis

- One tree for any comparison sorting on n-elements
- Each element comparison can be viewed as a splitting on a tree node.
- The tree contains all possible instruction traces.
- The running time of the algorithm == the length of path taken
- The best worst-case complexity == tree height

Lower-bound for Decision Tree Sorting

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

Proof. The tree must contain $\geq n!$ leaves, since there are $n!$ possible permutations. A height-h binary tree has $\leq 2^h$ leaves. Thus, $n! \leq 2^h$.

$$\begin{aligned}
\therefore\ h &\geq \lg(n!) && \text{(lg is mono. increasing)} \\
&\geq \boxed{\lg((n/e)^n)} && \text{(Stirling's formula)} \\
&= n \lg n - n \lg e \\
&= \Omega(n \lg n)\,. \quad \square
\end{aligned}$$

i.e."Stirling approximation":

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

More about Decision Tree

- 三个分支的决策树属于多叉树的一种，其构造速度要优于二叉树，但是精度比二叉树低.
- 决策树还用于模式分类(数挖)
  - http://www.hankcs.com/ml/decision-tree.html
  - http://www.chawenti.com/articles/18892.html