Preliminaries for Pruning

First, grow a vary large tree $T_{\max}$

1. until all terminal nodes are nearly pure;

2. or when the number of data in each terminal node is less than certain threshold;

3. or when the tree reaches certain size.

As long as the tree is sufficiently large, the size of the initial tree is not critical.

Notation : *subtree* $T' \prec T$, *branch* T_t.

Minimum Complexity-cost Pruning

For any subtree $T \prec T_{\max}$, define the Complexity-cost

$$R_\alpha(T) = R(T) + \alpha|T|, \tag{1}$$

- $R(T)$: RSS for regression tree T
- $|T|$: tree size, i.e., the number of leaf nodes
- $\alpha > 0$: cost (penalty) of adding a split

Questions: *i)* How to minimize (1) for a given α? *ii)* How to choose α?

Pick the best subtree that minimizes the cost

$$T(\alpha) = \text{argmin}_{T \preceq T_{\max}} R_\alpha(T) = \text{argmin}_{T \preceq T_{\max}} \Big[R(T) + \alpha |T| \Big]$$

$T(\alpha)$ may not be unique.

Define the optimal subtree $T^*(\alpha)$ to be the smallest one among $T(\alpha)$'s

(1) $R_\alpha(T^*(\alpha)) = \min_{T \preceq T_{\max}} R_\alpha(T)$.

(2) $T^*(\alpha) \preceq$ any $T(\alpha)$.

$T^*(\alpha)$ is unique.