The steps above generate a Solution Path:

$$T_{\text{max}} = T_0 \succ T^*(\alpha_1) \succ T^*(\alpha_2) \succ \cdots \succ \{\text{root node}\}$$

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots$$

All possible values of α are grouped into $(m+1)$ intervals:

$$\begin{aligned} I_0 &= [0, \alpha_1) \\ I_1 &= [\alpha_1, \alpha_2) \\ &\vdots \\ I_m &= [\alpha_m, \infty) \end{aligned}$$

where all $\alpha \in I_i$ share the same optimal subtree $T^*(\alpha_i)$.

Cross-validation

How to Choose α? K-fold Cross-validation (`rpart`):

1. Fit a big tree T_{max} and compute $I_0, I_1, \ldots, I_m$

$$\begin{aligned} \text{Set } \beta_0 &= 0 \\ \beta_1 &= \sqrt{\alpha_1 \alpha_2} \\ &\vdots \\ \beta_{m-1} &= \sqrt{\alpha_{m-1} \alpha_m} \\ \beta_m &= \infty \end{aligned}$$

where each β_j is a 'typical value' for its interval I_j.

2. Divide data into K groups and repeat $k = 1, \ldots, K$:

   - Fit a full model on the data set except the k-th group and determine the optimal subtrees:

$$T_0 \succ T^*(\beta_1) \succ \cdots \succ T^*(\beta_m) \succ \{\text{root node}\}$$

   - Compute the prediction error on the k-th group for each tree models.

3. Produce the CV plot over different α values and pick the optimal α_{min} or α_{1se}.