τ_B: Branch nodes

τ_L: Leaf nodes

L_t: The optimal misclassification loss in each leaf node t,

which is equal to the number of points in the node less the number of points of the most common label.

$\widehat{L}$: Simply predicting the most popular class for the entire dataset.

d_t: $\mathbf{1}$(node t applies a split) to track which branch nodes apply splits

$d_{p(t)}$: parents node

a_t: binary variables that sum to 1

α: complexity parameter

N_{kt}: to be the number of points of label k in node t

N_t: to be the total number of points in node t

N_{min}: the minimum leaf size

z_{it}: $\mathbf{1}$(xi is in node t) to track the points assigned to each leaf node

l_t: $\mathbf{1}$(leaf t contains any points) to enforce a minimum number of points at each leaf

n: number of training samples

Y_{ik}: an incorrect label prediction has cost 1, and a correct label prediction has cost 0, +1 if y_i = k, -1 otherwise

c_{kt}: binary variables to track the prediction of each node, where $c_{kt} = \mathbf{1}(c_t = k)$

ϵ_j: for feature j, the smallest non-zero distance between adjacent values of this feature

ϵ_{max}: maximum ϵ_j for all features