Post-Pruning Intuition

Monday 20 May 2024 3:13 PM

$R(T) = G.I$

Minimize this cost function during training

This can lead to complex tree having all leaves pure (overfitting)

→ We can add a penalty term

$$R_\alpha(T) = R(T) + \alpha |\tilde{T}|$$

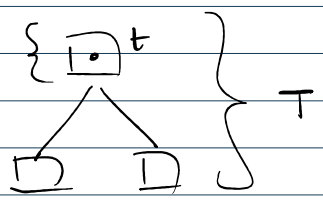
- α: Complexity parameter; controls the strength of the penalty.
- $R_\alpha(T)$: objective function
- $R(T)$: type of cost function i.e. (G.I)
- $|\tilde{T}|$: Type of penalty, but here it is the # of terminal nodes in T.

opposite forces

→ Now instead of D.T. having minimal $R(T)$, we are interested in D.T. " " $R_\alpha(T)$.

→ T with minimum $R_\alpha(T)$ compare to all other trees

→ w.r.t different values of α, different tree T will have minimal $R_\alpha(T)$ we will get another T.



$R(t) > R(T)$ i.e. G.I(parent) > G.I(childs)

$$R_\alpha(t) = R(t) + \alpha(1)$$

$$R_\alpha(T) = R(T) + \alpha|\tilde{T}|$$

At which value of α, $R_\alpha(T) = R_\alpha(t)$

$$R(T) + \alpha|\tilde{T}| = R(t) + \alpha$$

$$\alpha|\tilde{T}| - \alpha = R(t) - R(T)$$

$$\boxed{\alpha = \frac{R(t) - R(T)}{|\tilde{T}| - 1}}$$

$\uparrow \alpha$ — less decrease in impurities when we split parent into childs.

$\uparrow \alpha$ — high complexity added to the DT when parent is splitted into childs

→ can lead to overfitting [weakest link]