“Examples of Post-Pruning” Potential Error

Description: In the slide shown below, the authors state that when using Optimistic Error one should not prune in either case #1 or case #2. However, in case #2, there is no error improvement when using optimistic error; hence, by Occam’s Razor, the simpler solution should be preferred (i.e. to prune).

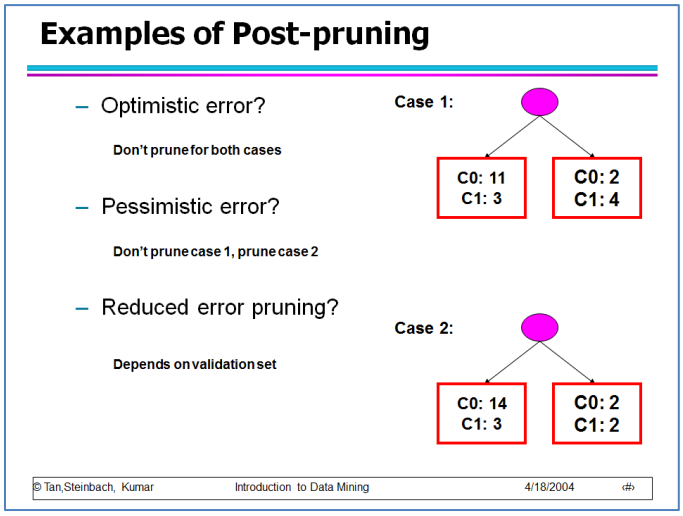

Figure 1 – “Examples of Post-Pruning” Slide from Tan et. al.

Proof: Below is case #2 redrawn with the parent node element shown more explicitly.

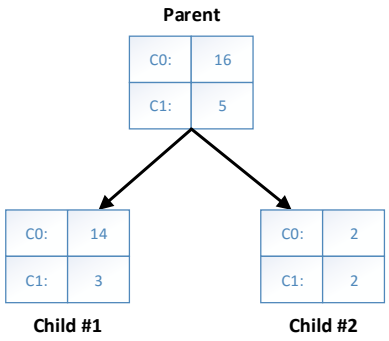

Figure 2 – Case #2 Redrawn with the Parent Node

The training error for the parent node is:

$$e(Parent) = \frac{\min(5,16)}{21} = \frac{5}{21}$$

Similarly, the training error for the child nodes is:

$$e(Child) = \frac{\min(3,14) + \min(2,2)}{(3+14)+(2+2)} = \frac{(2+2)}{17+4} = \frac{5}{21}$$

When using Optimistic Estimation, there is not a change in the error between the parent and child nodes. Hence, splitting the parent node complicates the model without improving its performance. Per the lecture slides, Occam’s Razor stipulates, “Given two models of similar generalization errors, one should prefer the simpler model over the more complex model.” Clearly, in this case, with and without pruning have similar (i.e. identical) generalization errors when using optimistic estimation. Hence, one should choose to prune since that is the “simpler model.” This is different than what is specified on the slides, and I argue the slide is in error.