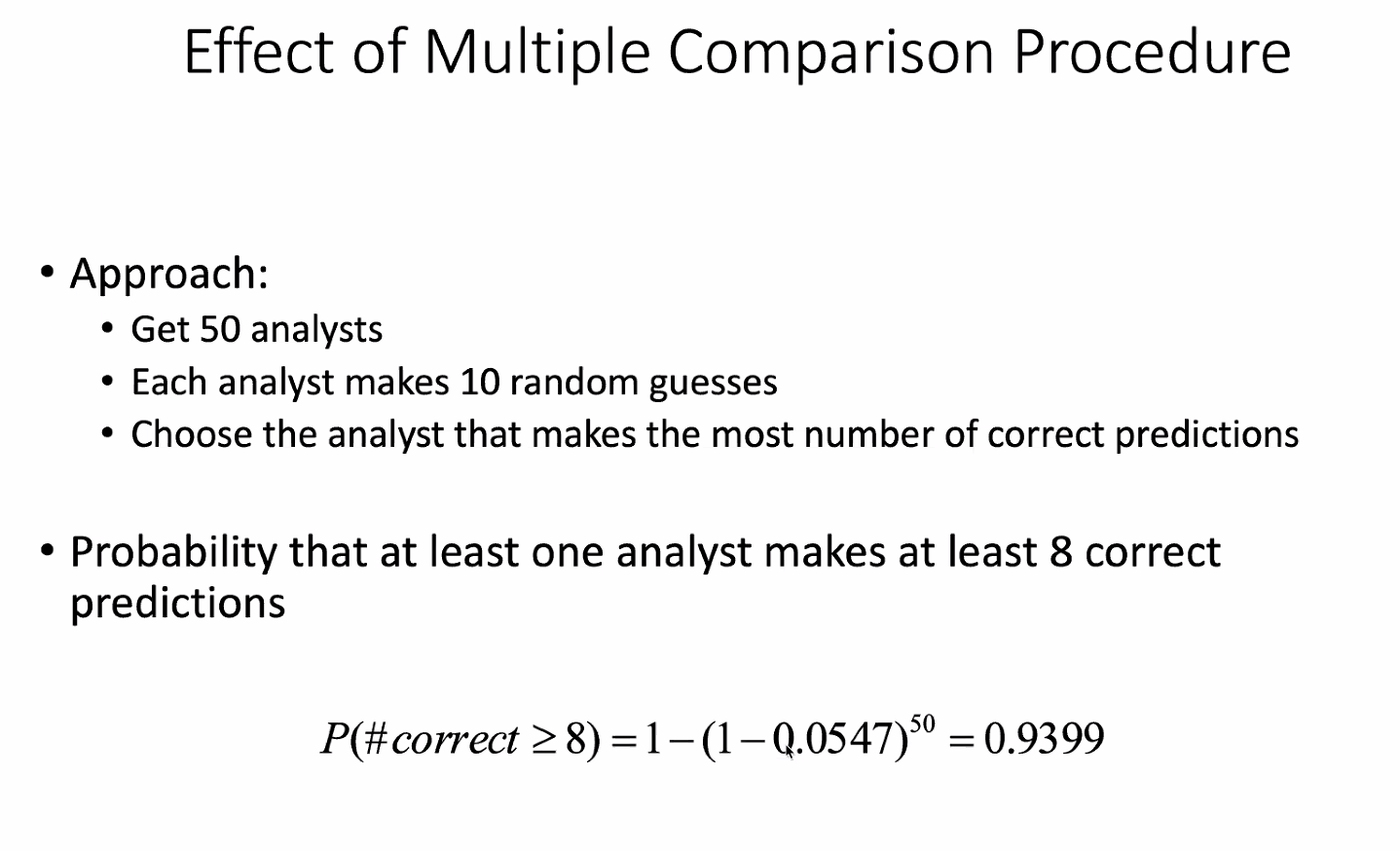
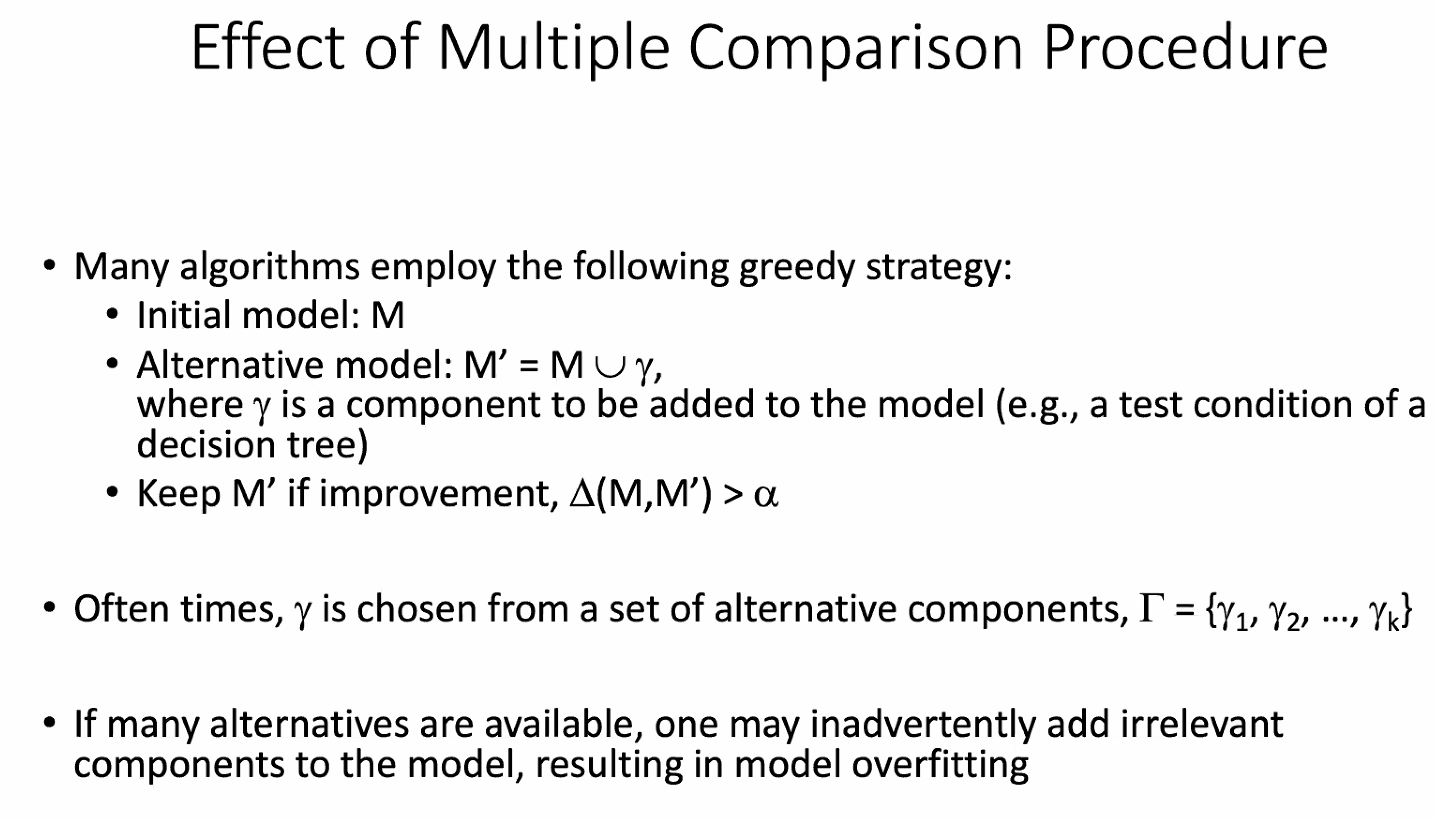
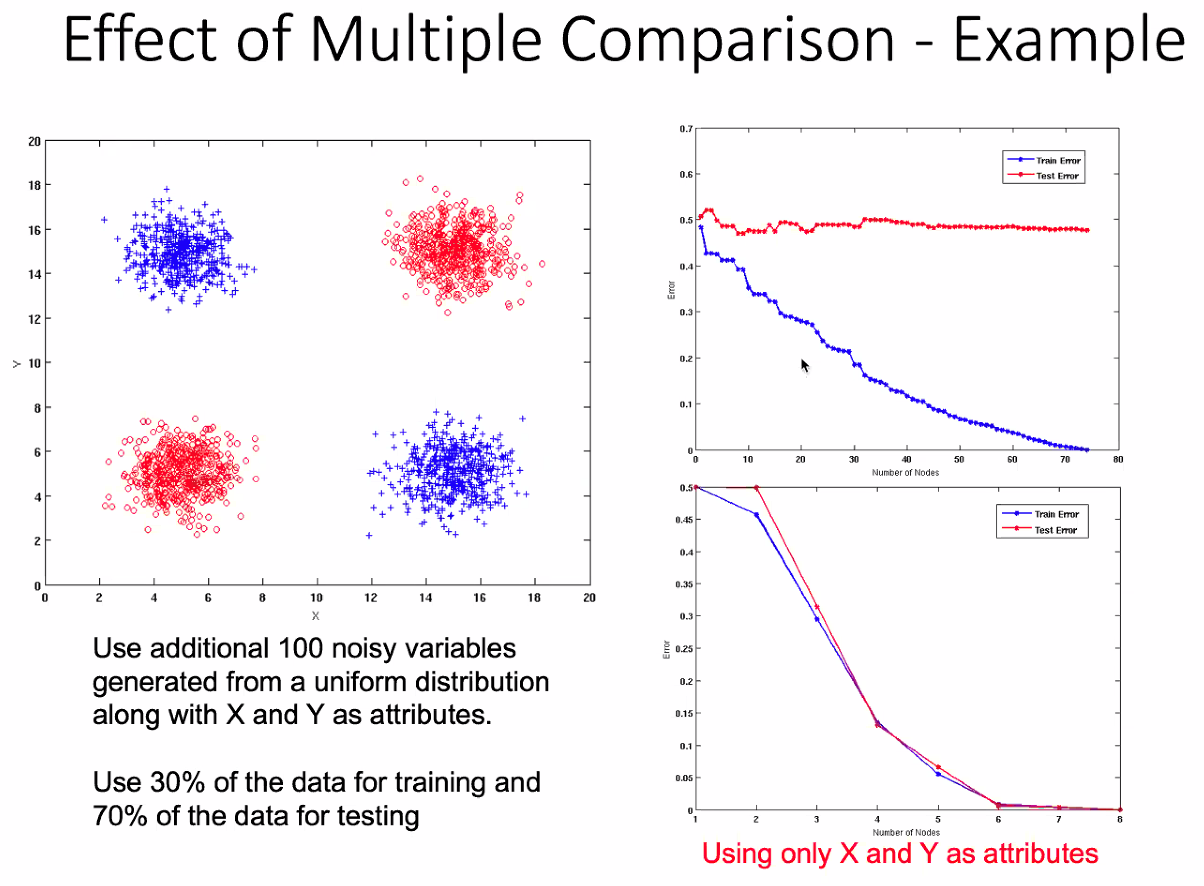


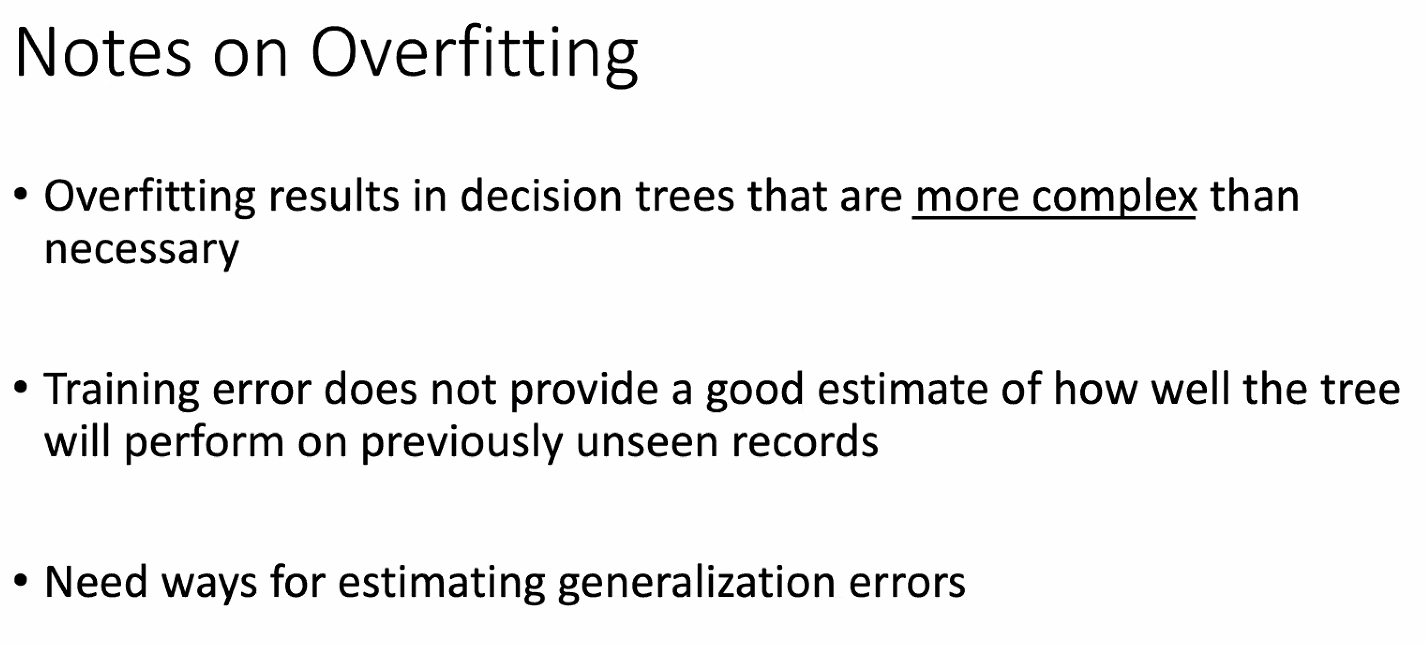
Right is better more data less noise, less error😊





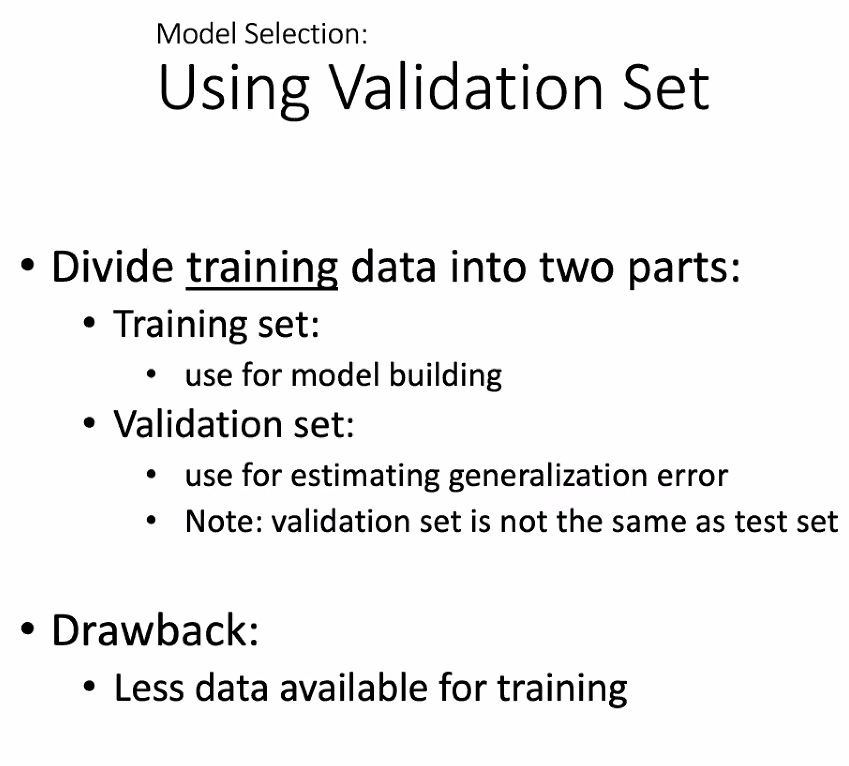


The noise is used has higher information gain x and y uninformative on their own, but together good!

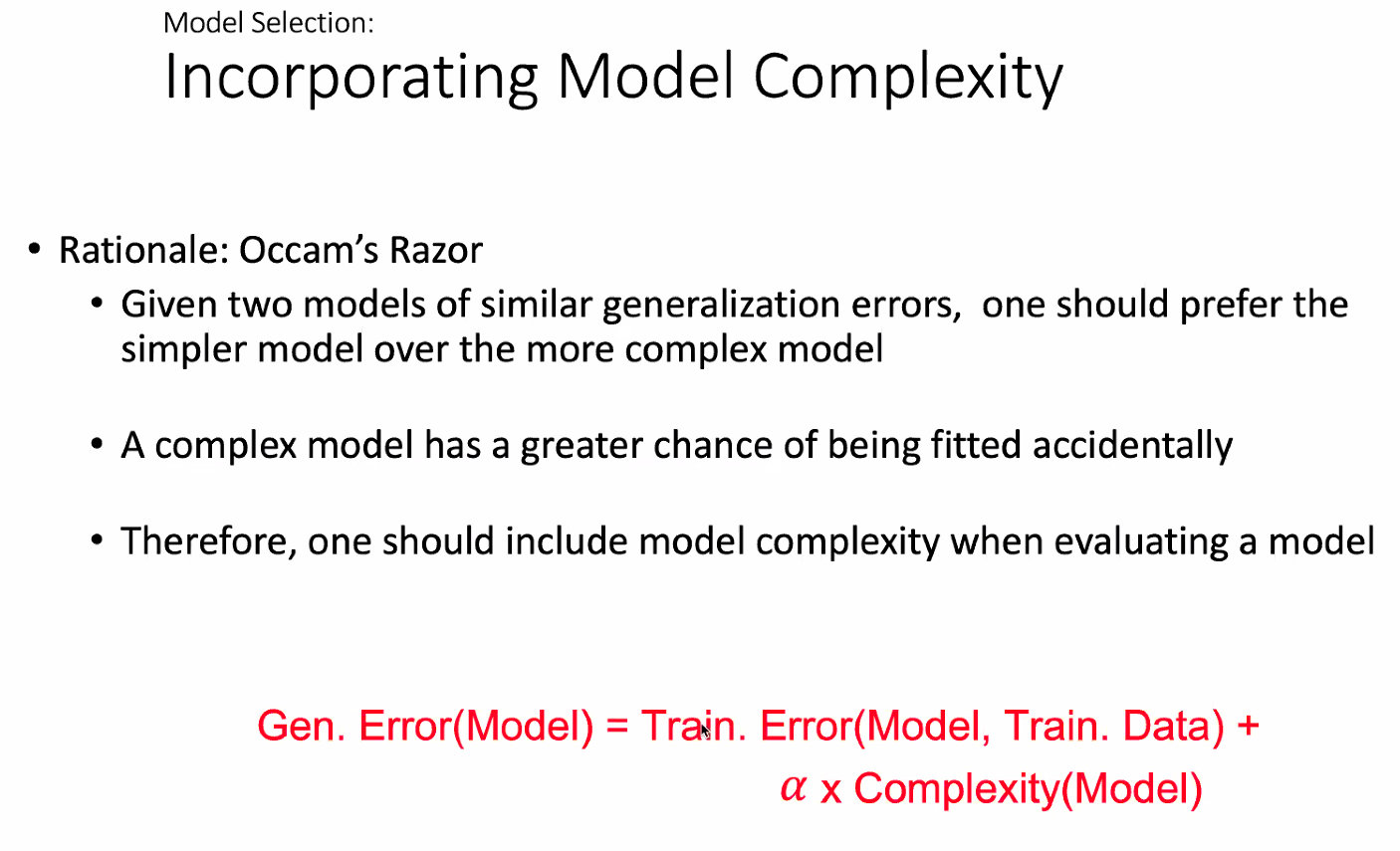


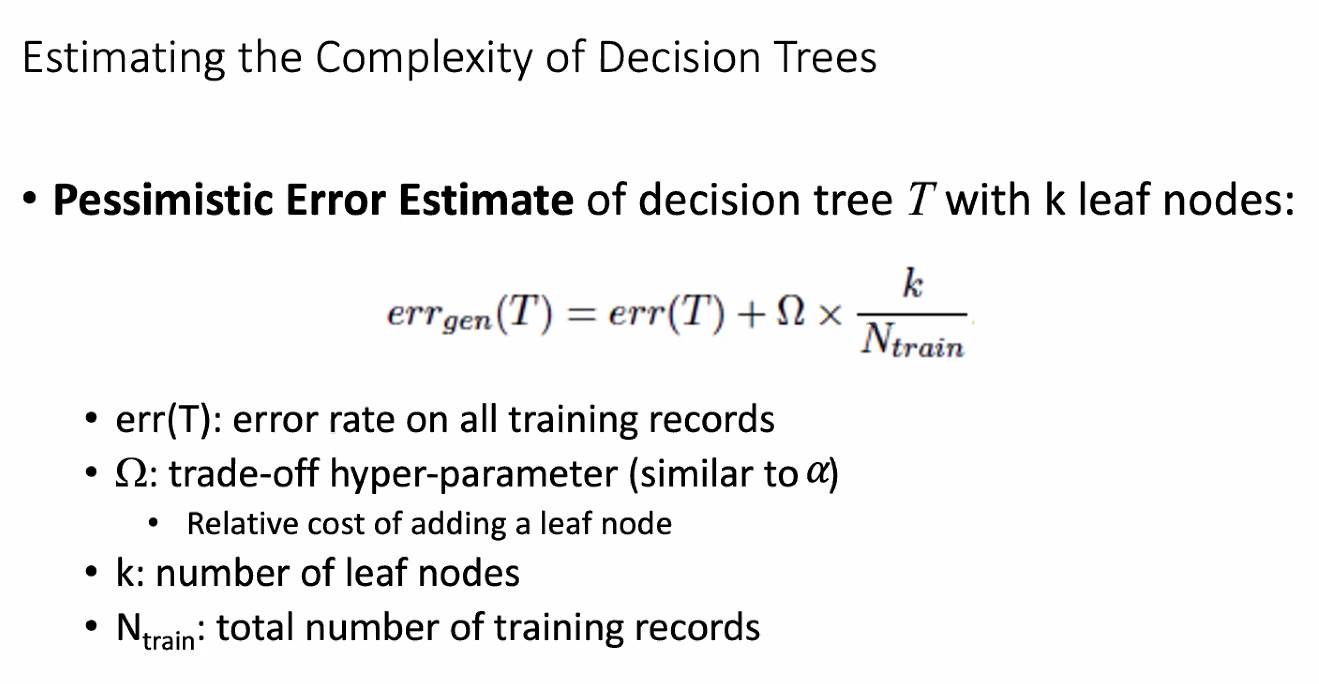
Training error: not a good metric for decision tree

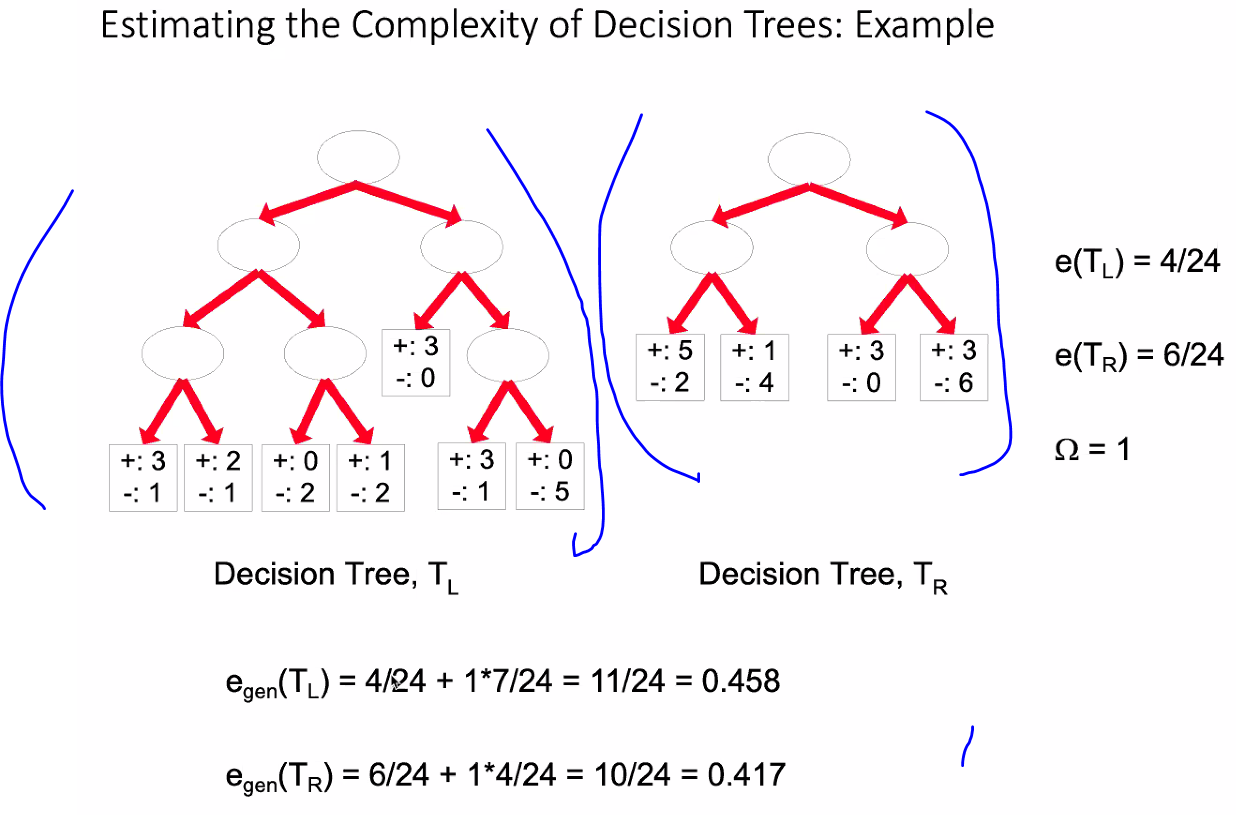
USE a validation set!



For model selection decision trees:







Increase complexity, decrease training error trade-off

