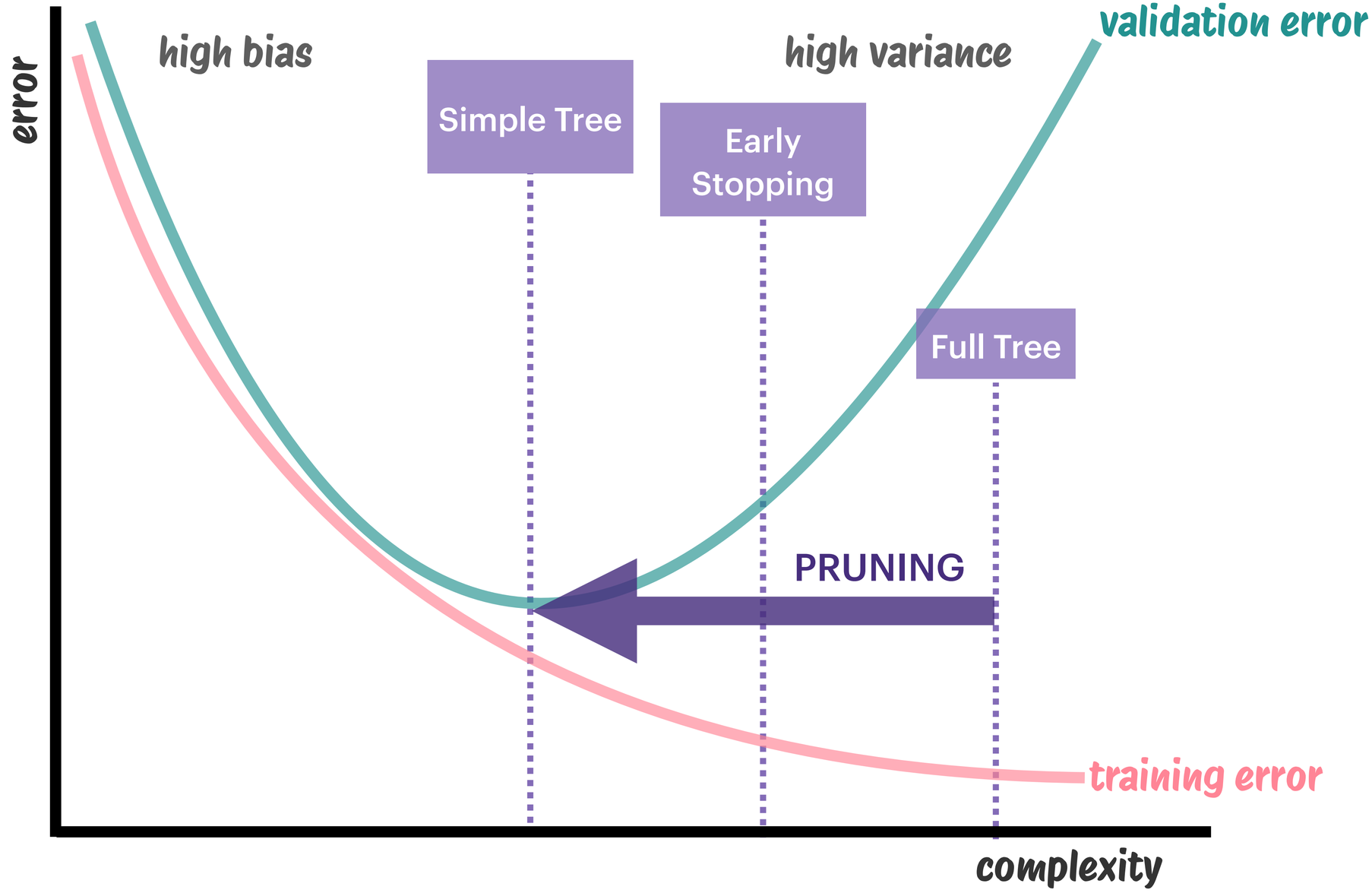
# Motivation for Pruning



# Cost Complexity Pruning

- we can obtain a simpler tree by ‘pruning’ a complex one
- we select from an array of smaller subtrees of the full model that optimizes a balance of performance and efficiency

$$C(T) = \text{Error}(T) + \alpha |T|$$

where $T$ is a decision subtree

$|T|$ is the number of leaves in the tree

$\alpha$ penalizes model complexity

1. Fix $\alpha$
2. Find best tree for a given $\alpha$ and based on complexity $C$
3. Find the best using CV and error measure