# Pruning

Saturday 4 May 2024 12:56 PM

→ The reason for DT's overfitting problem lies in it's stopping criteria

↳ Achieve total purity at each leaf node.

→ DT's are susceptible to overfitting b/c they can potentially create very complex trees that perfectly classify the training data but fail to generalize to new data.

→ Pruning helps to solve this issue by reducing the complexity of the decision tree, thereby improving its predictive power on unseen data.

→ Pruning types
- Pre-pruning (Early stopping)
- Post-pruning (cost complexity pruning)

→ A seemingly worthless split early on in the tree might be followed by a very good split, a split that leads to a large reduction in impurity index later on. [pre pruning leads to short-sightedness] (early stopping)

→ A better approach might be to let the tree grow to max depth & prune it back in order to obtain a subtree [post-pruning]