

- Decision trees are highly prone to overfitting.
- As you increase the depth, you increase the complexity of the decision boundaries.
- How do we pick simpler trees?
 - 1) Early Stopping: Stopping learning algorithm before tree became too complex.
 - 2) Pruning: Simplify tree after learning algorithm terminates.
- Early stopping condition 1: Limit the depth of a tree.
 - However, if you limit the depth, you're cutting every branch the same way, but some branches might want to go further than others.
- Early stopping condition 2: Use classification error to limit depth of tree.
- Early stopping condition 3: Stop if number of data points contained in a node is too small.
- Handling missing data Strategy 1: Purification by skipping
 - Idea 1: Skip data points where any feature contains a missing value.
Note: Make sure only a few data points are skipped.
 - Idea 2: Skip an entire feature if it's missing for many data points.
Note: Make sure only a few features are skipped.
 - Pros and Cons:

Pros: 1) Easy to understand and implement. 2) Can be applied to any model (decision trees, logistic regression, linear regression...)

Cons: 1) Removing data points and features may remove important information from data. 2) Unclear when it's better to remove data points versus features. 3) Doesn't help if data is missing at prediction time.
- Handling missing data Strategy 2: Purification by imputing
 - Impute each feature with missing values: 1) Categorical features use mode: Most popular value (mode) of non-missing x_i ; 2) Numerical features use average or median: Average or median value of non-missing x_i .
 - Pros and Cons:

Pros: 1) Easy to understand and implement. 2) Can be applied to any model (decision trees, logistic regression, linear regression...). 3) Can be used at prediction time; use same imputation rules.

Cons: May result in systematic errors.
- Handling missing data Strategy 3: Adapt learning algorithms to be robust to missing values
 - Pros and Cons:

Pros: 1) Addresses training and predictions time. 2) More accurate predictions.

Cons: Requires modification of learning algorithm.