

# Ensemble Learning For Decision Trees



- 
1. Random Forests
  2. Boosting

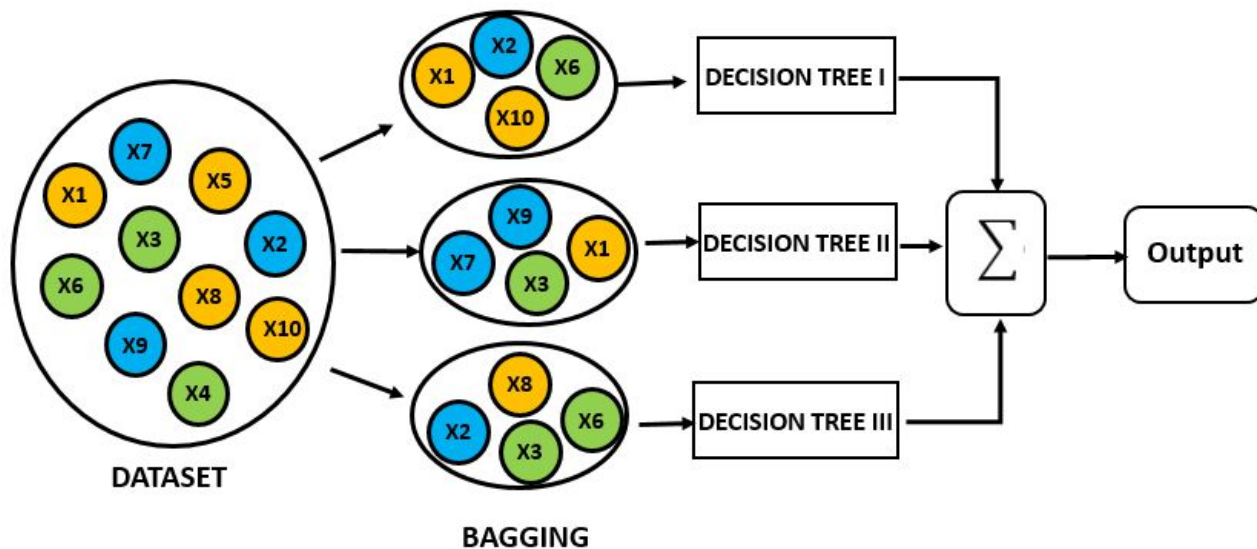


# Random Forests



# How Random Forests Work

- A random forest is an ensemble of decision trees, each of which are trained on random subsets of the data



# Bagging vs Pasting

- When we sample with replacement, the method is called bagging
- When we sample without replacement, the method is called pasting



# Random Subspaces

- We can also take a random subset of the features when training individual decision trees
- This makes training the decision trees faster



# Boosting



# What Is Boosting?

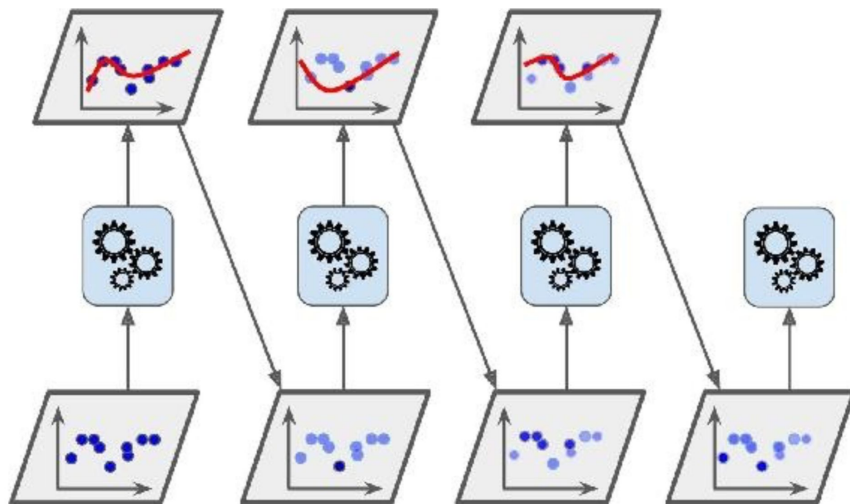
- Boosting refers to an ensemble method that combines a bunch of weak learners into a strong learner
- We train the predictors sequentially, each trying to correct its predecessor





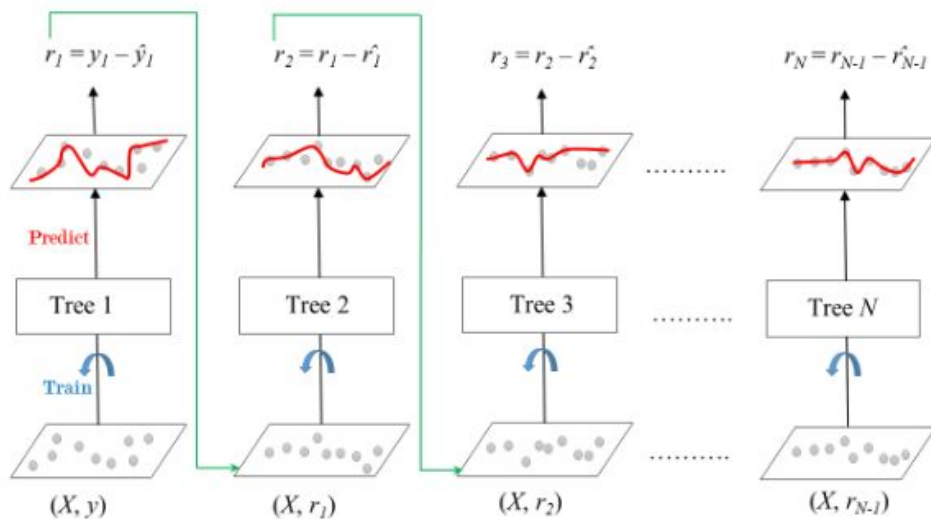
# AdaBoost

- In AdaBoost, we pay more attention to the training instances that the predecessor underfitted



# Gradient Boosting

- Gradient Boosting works by fitting a model to the residual error of the previous predictor



# Questions to Answer

1. When we use a random forest with bagging, there will be some data points that are never used to train any of the decision trees (why is this the case?). What can you do to alleviate this problem?
2. Which ensemble methods can easily be parallelized across multiple CPUs?
3. When should we use hard voting vs soft voting? Or is this something that is heuristic and we just have to try both possibilities and see which one is better?
4. When training the individual trees, we take a different sample from the training data for each tree. Why do we do this, instead of having each tree train on the same full data?

