

Random Forest

Xi Han

Haoran Zhao

Random Forests

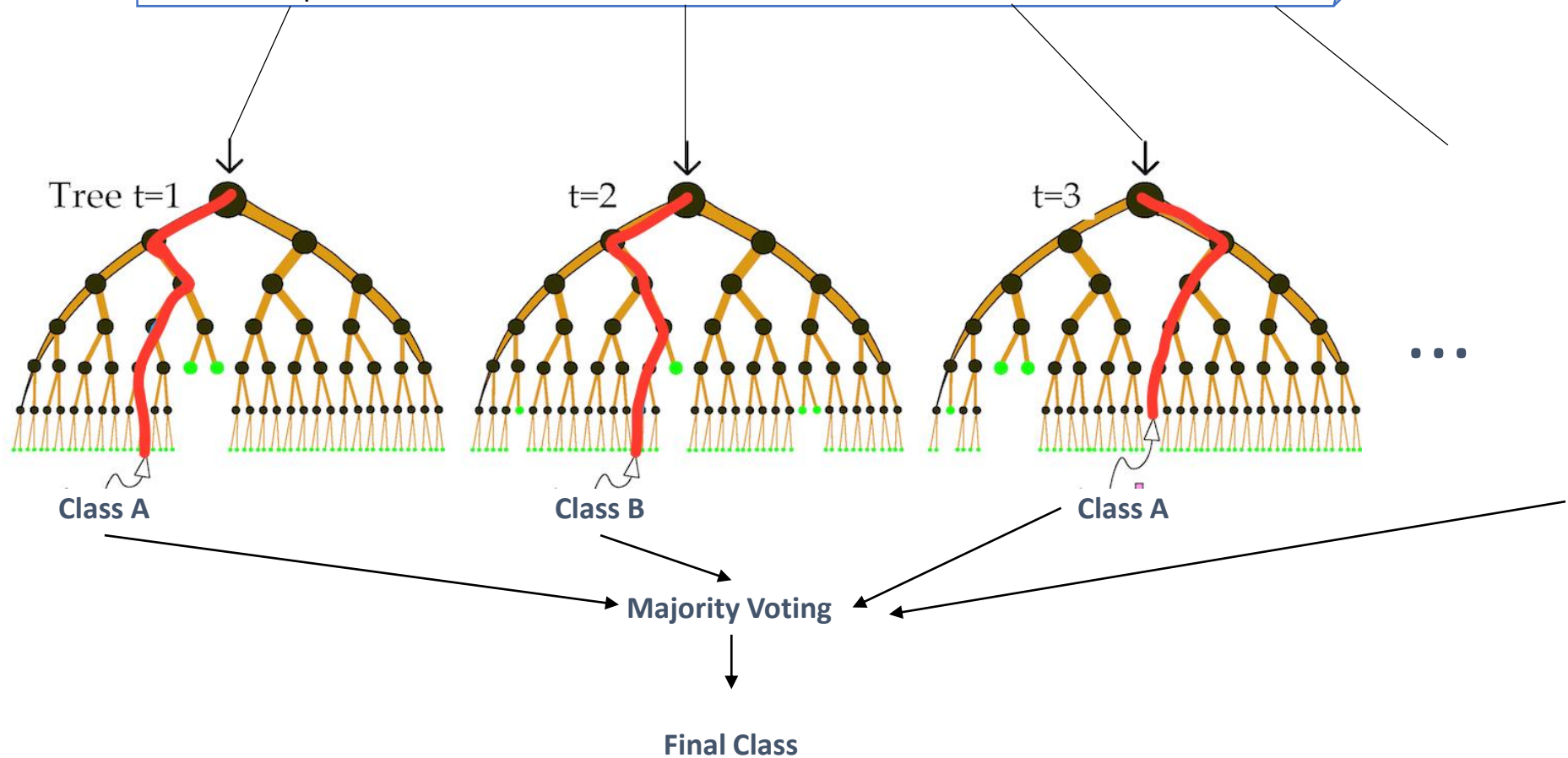
Decision Tree is one in which the final outcome of the model is based on a series of comparisons of the values of predictors against threshold values

Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone

Random Forests is an ensemble of independent Decision Trees

How does Random Forests work?

1. **Samples Bootstrapping** : random subset of samples with replacement for each tree
2. **Predictors**: for each tree, at each split, **randomly** select a set of predictors from the full set of predictors



Advantage and Disadvantage of Random Forest

- Advantage:
 - The model outputs **importance of variable**.
 - Non-linear relationship between predictors and target variable
 - Harder to overfit and less variance compared to decision tree
- Disadvantage:
 - Hard to interpret
 - Could be memory intensive and longer to compute for large number of trees with many depths