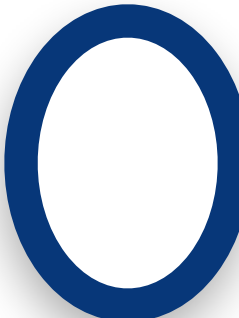

H2O.ai

RF Tree Variance: Splits

2) Do not consider all of the features for each split

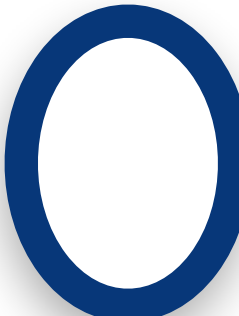
- default is often
- also speeds up computational time

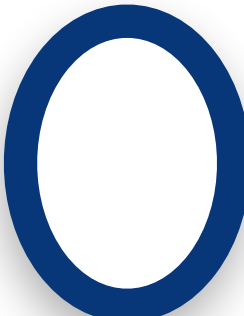


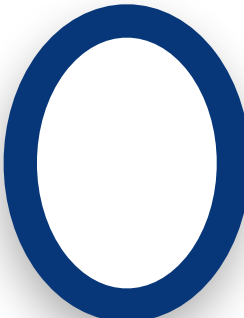
Example:

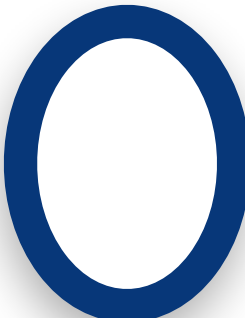
features = {age, sex, income, education}

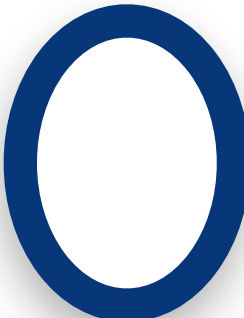
Split 1:
best split from {age,income}

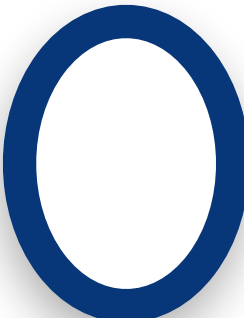






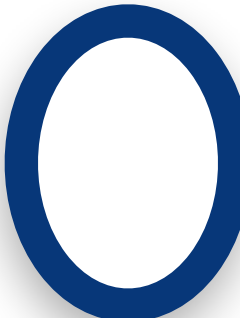


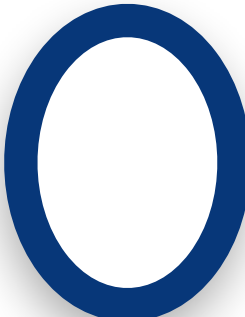




Split 2b:
best split from {sex,income}

Split 2a:
best split from {sex,education}

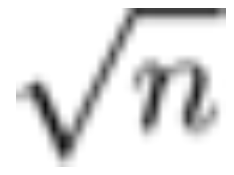




Split 3a:
best split from {sex,education}

Additional methods for variance:

- do not consider all features for each *tree*
- H2O: random histograms



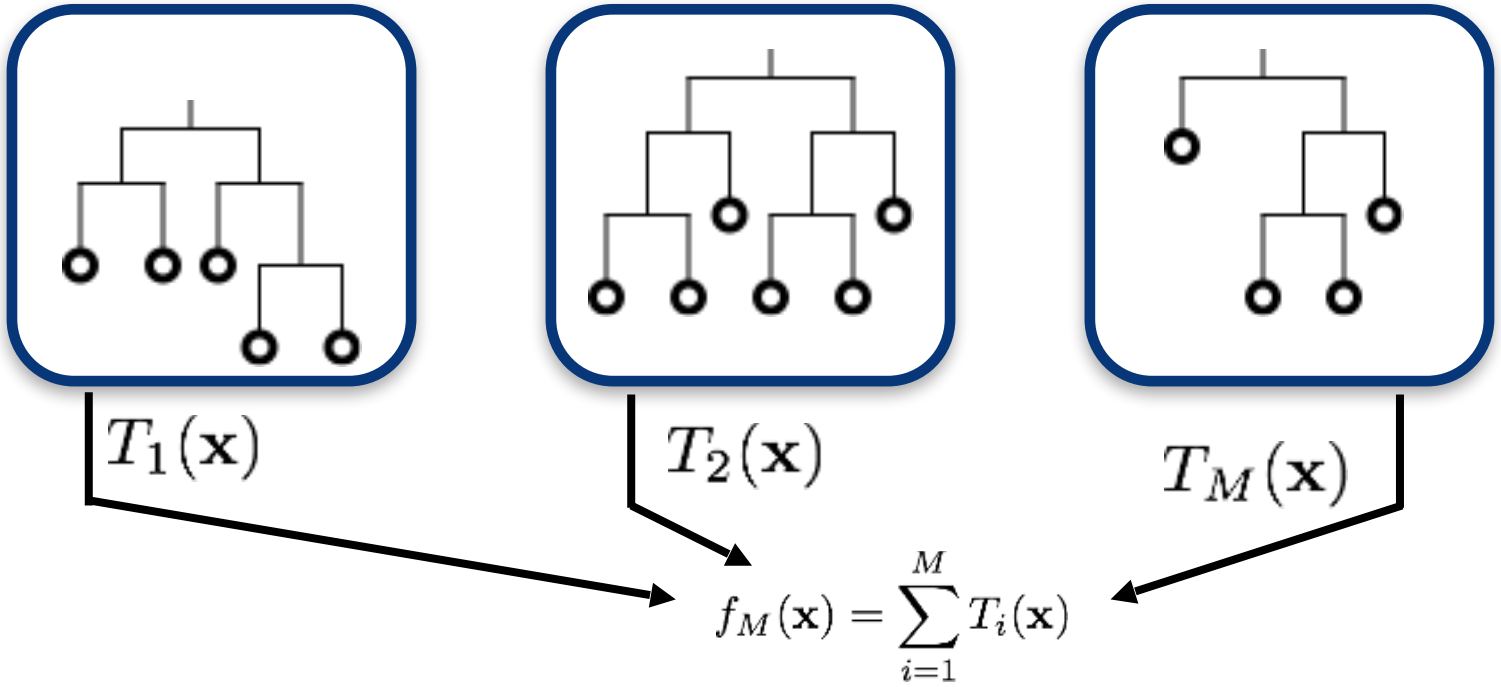
Distributed RFs

Method 1: Parallelize by tree

✔ low communication between nodes (only to reduce as part of map/reduce)

✗ every node must have all data in memory

✗ does not generalize to GBMs



RF Tree Variance: Splits

2) Do not consider all of the features for each split

- default is often $\sqrt{n}$
- also speeds up computational time

Additional methods for variance:

- do not consider all features for each *tree*
- H2O: random histograms