



Distributid RRFs in H2O

Method 2: Parallelize by data







H₂O.ai

H₂O.ai













- ✓ can handle much larger data sets
- ✓ sequential training generalizes to GBMs
- ✗ more communication between nodes to reduce the computations
- ✗ relies on approximate split calculation via histogram









































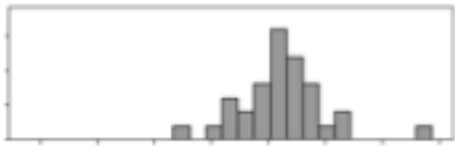
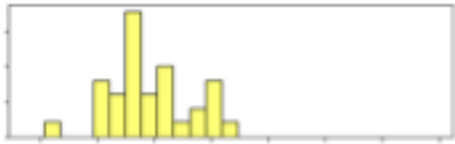


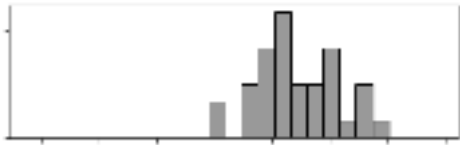
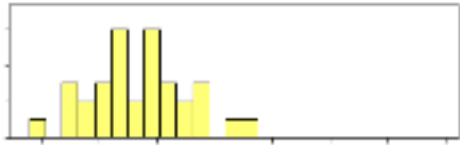


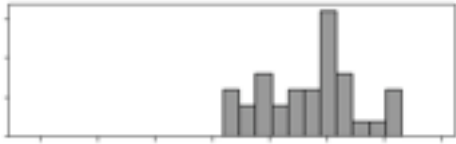
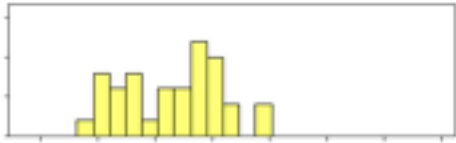


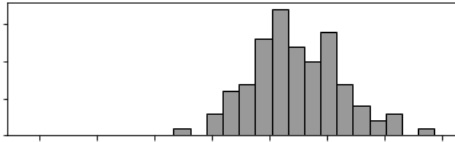
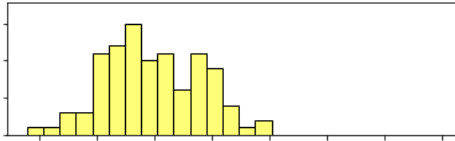












RF Parameters: Setting Up a Model

Python API

```
h2o.estimators.random_forest.H2ORandomForestEstimator  
- h2o.estimators.estimator_base.H2OEstimator  
-- h2o.model.model_base.ModelBase
```

model_id: specify a custom name for the model to use as a reference. (default: randomly generated)

checkpoint: load a previously generated model

seed: specify a pseudorandom seed, for reproducibility

verbose: print verbose scoring history to console

Distributed RFs in H2O

Method 2: Parallelize by data

