Random Forest

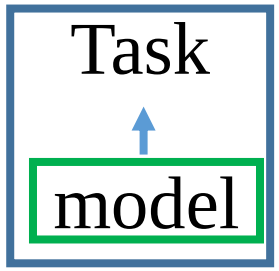
Liang Liang

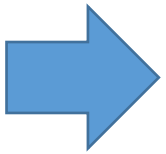
Combination of Models

- There is no algorithm/model that is always the most accurate
- We can build many models (e.g., simple classifiers or regressors) and combine them into a single "strong" model
- Different models may use different

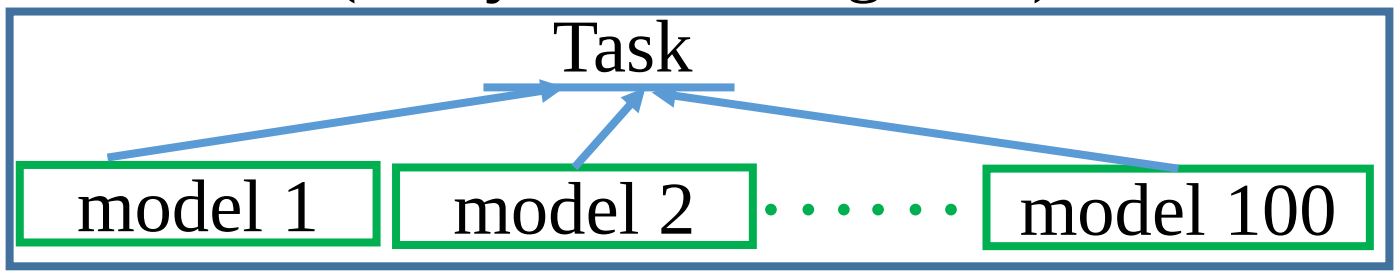
  learning algorithms, hyper-parameters, training sets, model type/structure, etc

a single model



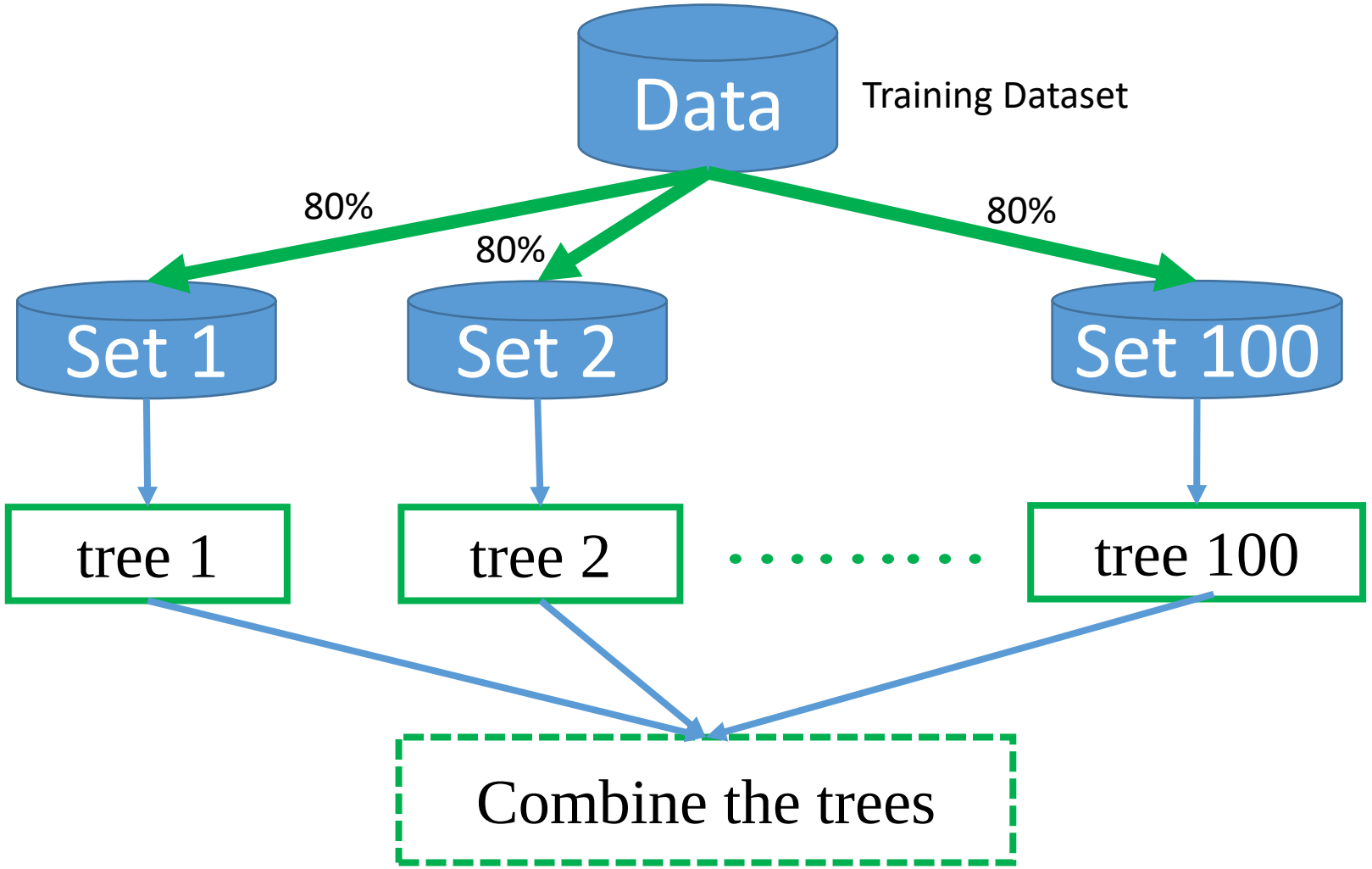


Ensemble (many models, e.g.,100)



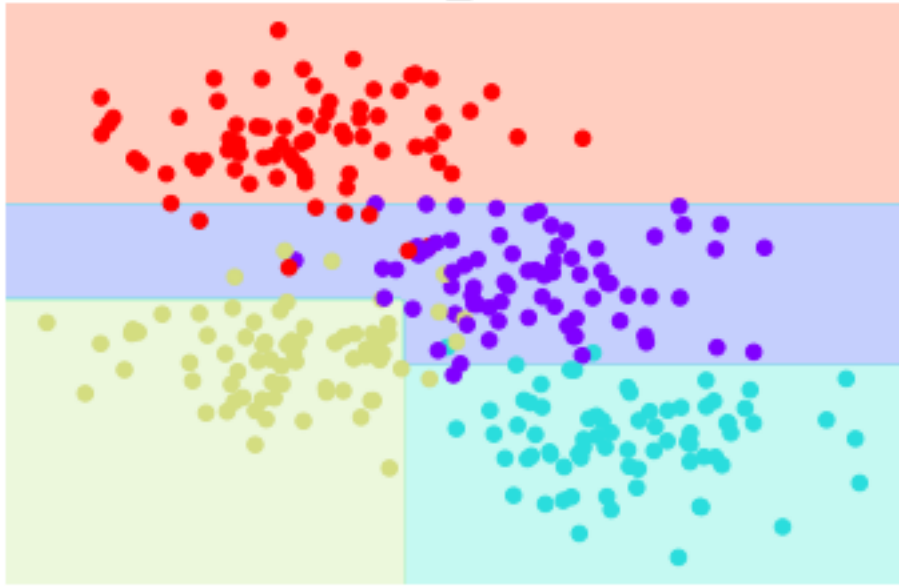
Random Forest (a bag of trees)

- A random forest (RF) is a combination of many decision trees.
- Each tree is trained on a randomly selected subset of the training data
- The output of a RF is average (regression) of the individual trees.
  or majority vote (classification) of the individual trees.

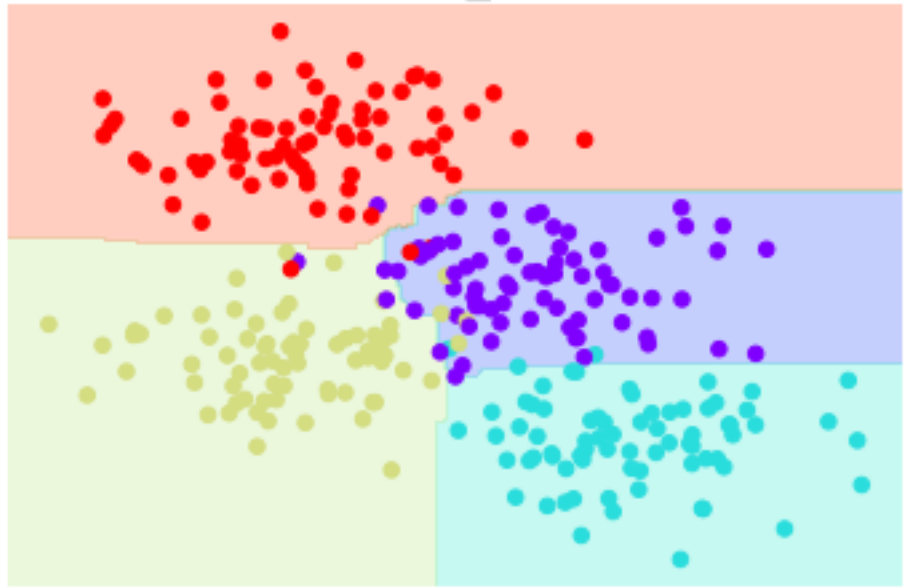

Data
Training Dataset
80%
80%
80%
Set 1
Set 2
Set 100
tree 1
tree 2
tree 100
Combine the trees

see the demo:
Decision_Trees_and_Random_Forests_basics.ipynb

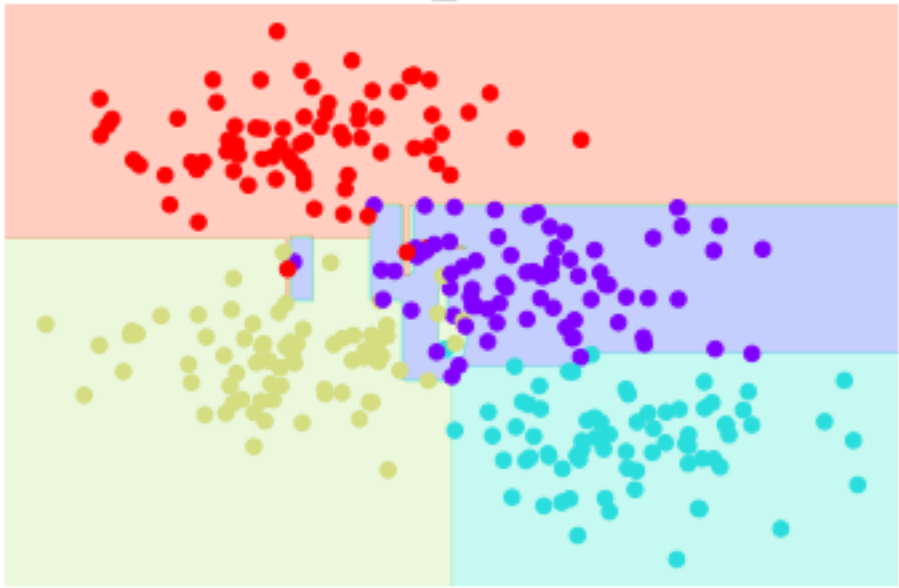
tree max_depth=3



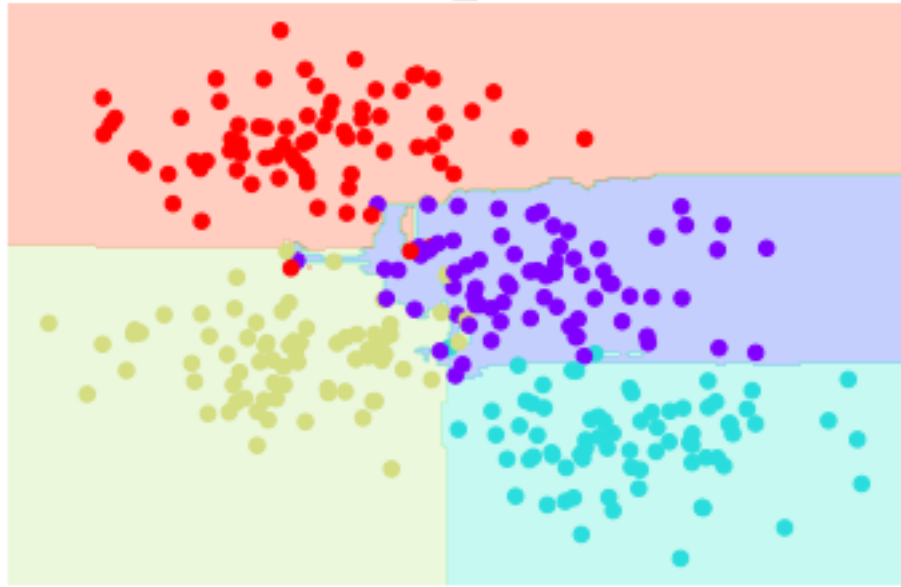
forest max_depth=3

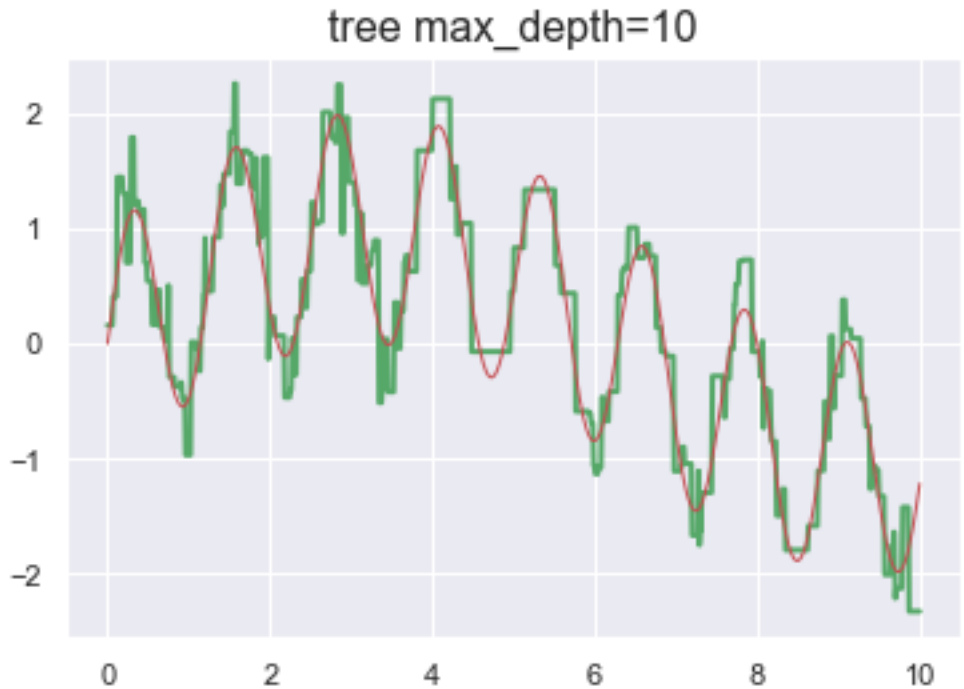


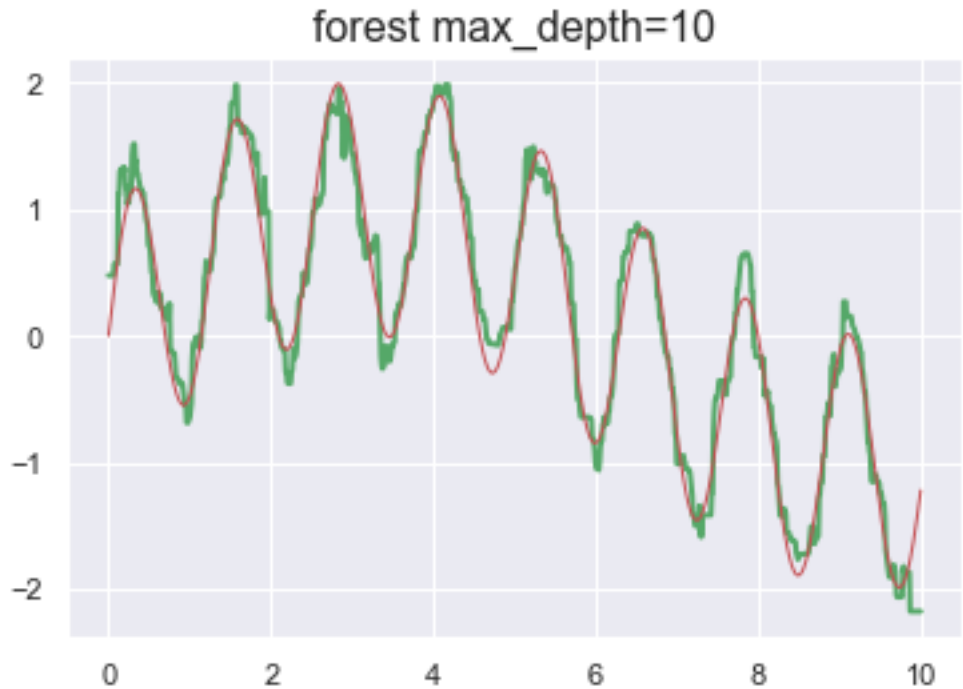
tree max_depth=10



forest max_depth=10




tree max_depth=10
2
1
0
−1
−2
0
2
4
6
8
10


forest max_depth=10
2
1
0
−1
−2
0
2
4
6
8
10