

Decision Tree

- Since single tree is prone to model overfitting, we use ensemble tree pruning techniques (bagging, boosting). Decision overfitting is not a problem, but ensemble effect can improve accuracy and reduce variance.

(i) pre pruning

(ii) post pruning

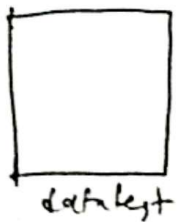
Ensemble Methods

↳ Random forest

Forest = multiple trees = 1

↳ Bagging → parallel
↳ Boosting → sequential

Random forest → By default 100 trees.



→ Best learning model

100 tree subset

new feature sampling

↳ bagging / boosting

unique subset

parameter

• hyperparameter

→ Grid search cv

→ Random cv

↳ no. of iteration
random forest

best parameters - those that give the best accuracy
train accuracy

↳ best accuracy

hyperparameter tuning