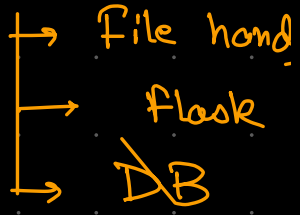
Welcome back

1. Pruning DT
2. Bagging & Boosting
   - Bagging ⇓ RF
   - Boosting ⇓ Adaboost
3. SVM/SVR *
4. Unsupervised

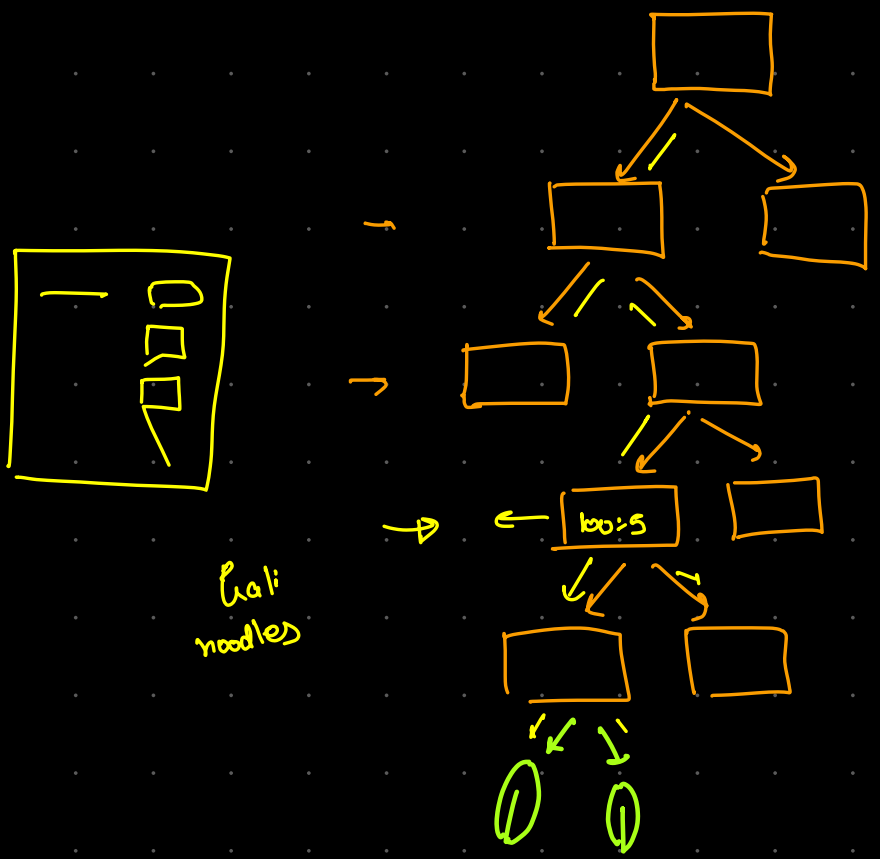


← End to end to Project

Future classes in morning from 10-1 pm.

Sat - Sun

Limitation in DT



So our DT has more chance of overfitting

⇓

low bias

high variance

Decision tree with complete depth is more prone to overfitting.

⇓

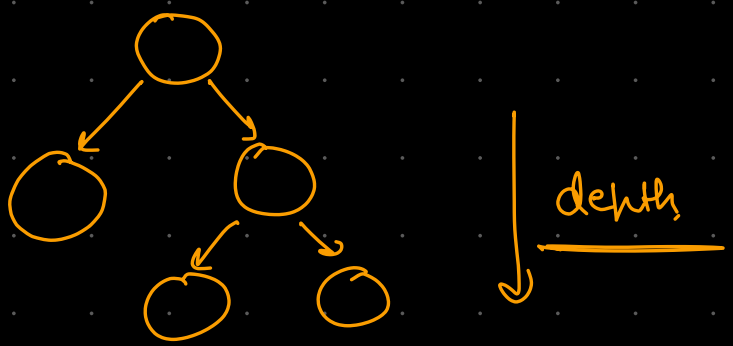
Pruning ← cutting of your tree.

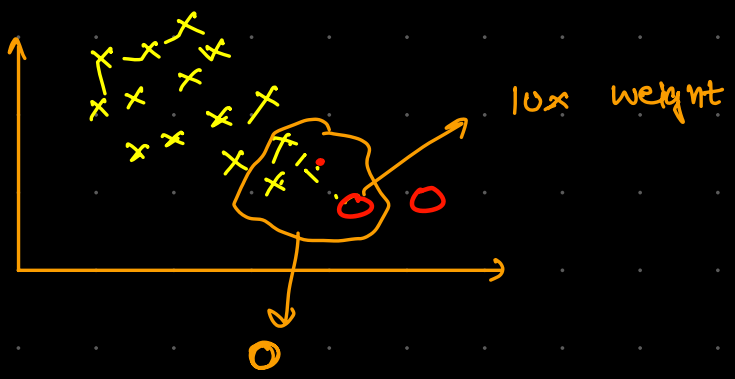
Post Pruning

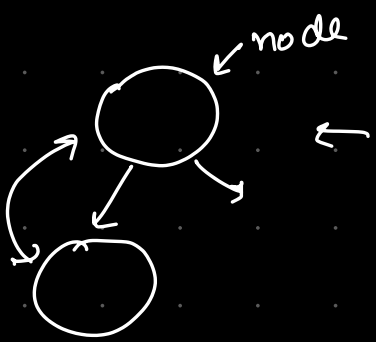
cut after it has grown to full length.

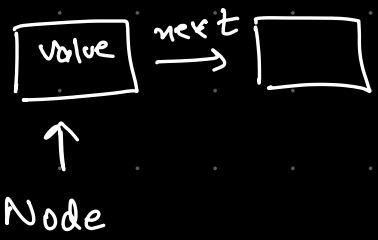
Pre pruning

Monitoring & cutting while/before getting built.









Node

Random Forest

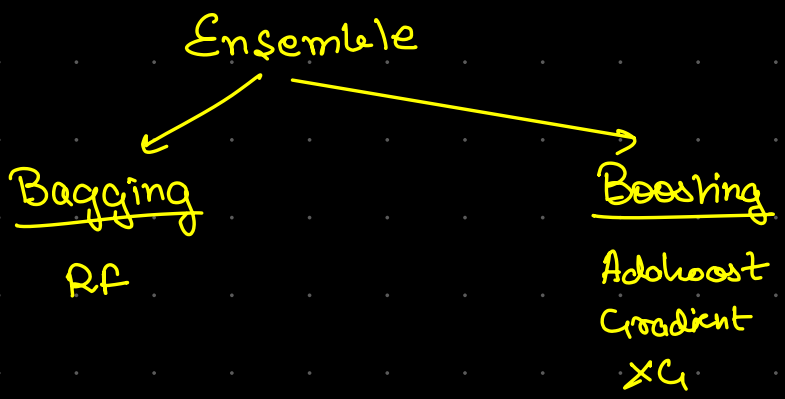
↳ we will make lots of trees.

DT we know that it is prone to Overfitting.

OF → low bias
high variance

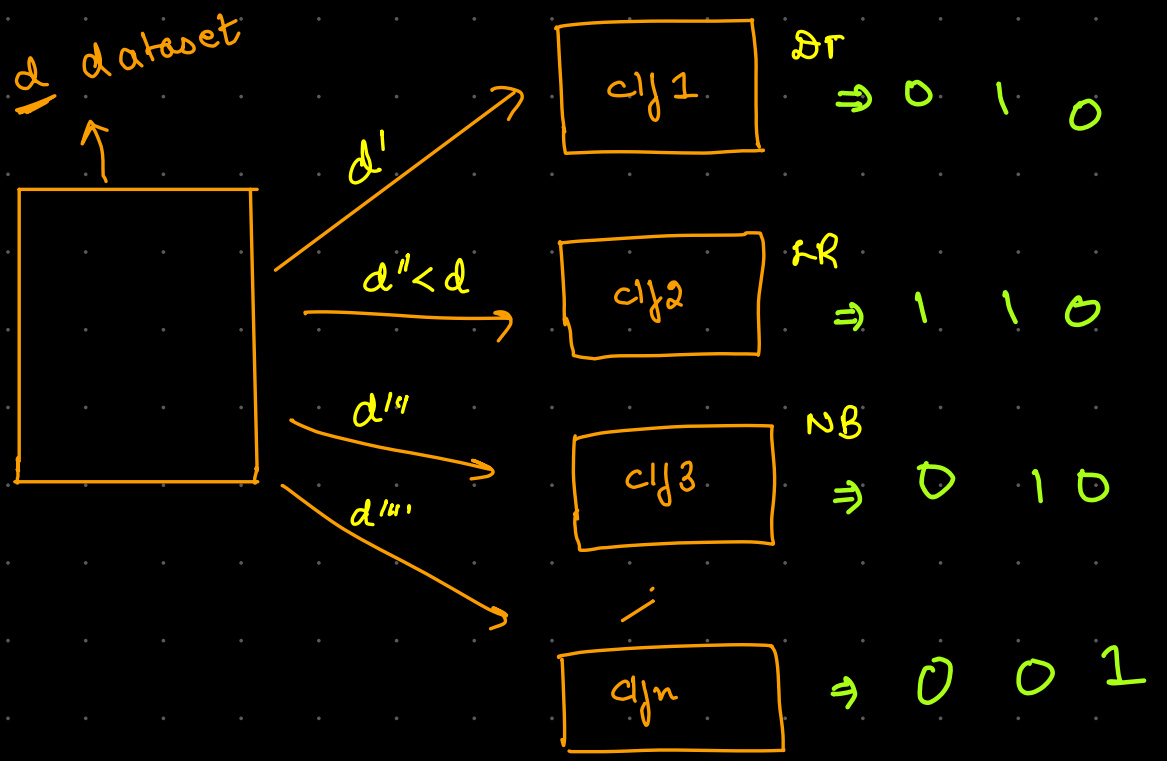
We will try to move from high var → low var

For the same we use
ensemble technique



Instead of trying a single algo on our dataset, we will try to use multiple algorithm & calculate the output

Bagging

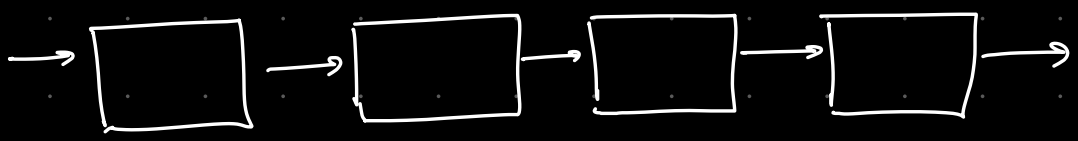


Classification ⇒ majority

regression ⇒ mean

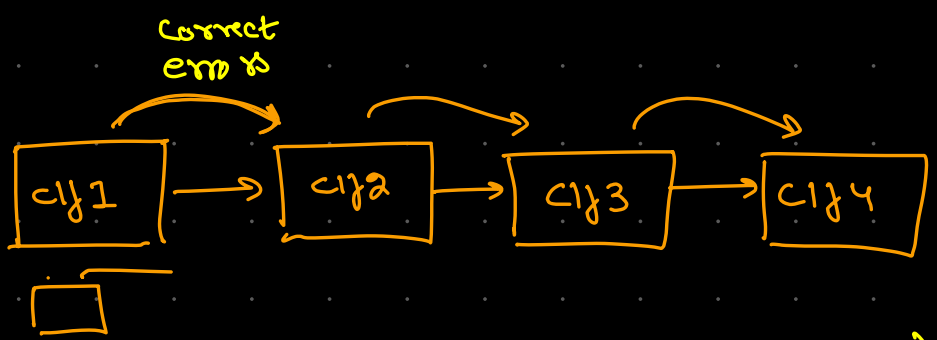
Majority voting Classifier

Boosting

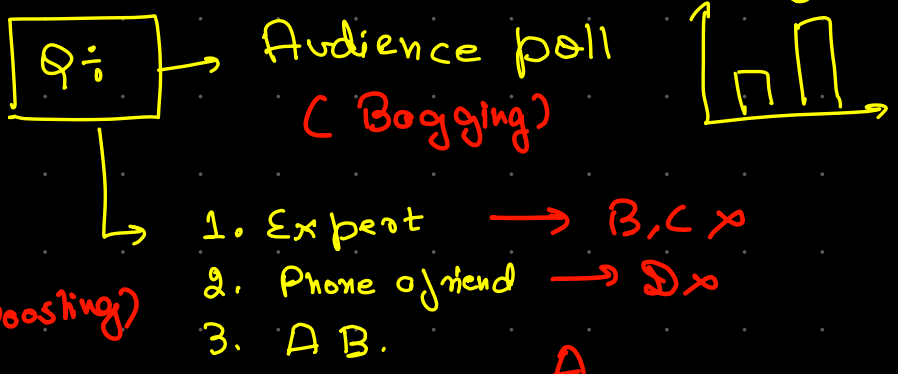


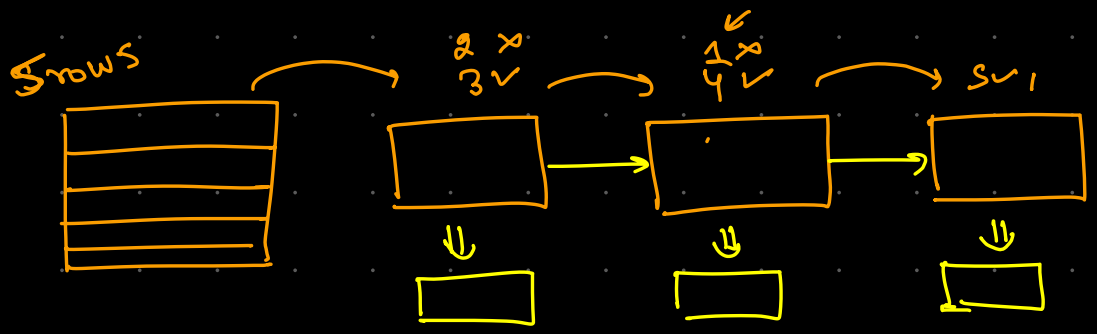
Boosting your learning

Weak learner ⇒ Each individual algorithm



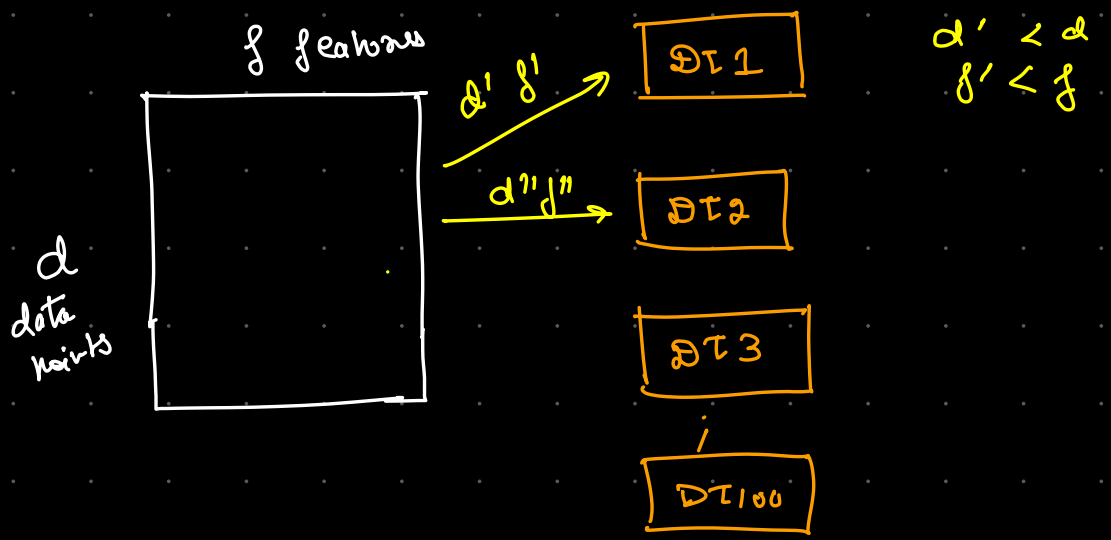
KBC




Srows
2
3
7
4
Sw1

Random Forest

Classifier and a regressor



d' = subset of rows
f' = subset of features
} duplicates can be there

Providing subset of features and data helps one in avoiding bias & also to not focus on something specific.

Adaboost [Boosting Algo]