Hierarchical algos ⇒ Tree } Information gain
non-parametric } GINI INDEX

Entropy,
$p(x) \log p(x)$

Draw Backs ?

in DTs :

Inference on

1. Overfitting → Out of Bag Samples (OOB) ⇒ DT fails.
2. Continuous
3. FFs not suitable for large

Reasons DT as weak classifier.

Only X Unsupervised.

Parameters
< m, c → data [illegible]

1) Hyperparameters
max_depth

[illegible] (L1 → L2)

Slope

Gradient descent

Random Forest

I am looking at handling Dis adv of DT.

↳ ISSues OOB's

Random ⟷ Forest

→ Trees more Trees.

Random

Random collection of Trees!!! ⇒ Decision Trees.

ENSEMBLE LEARNING } ⇒ Combining Multiple learners together

↓

ENSEMBLING Techs

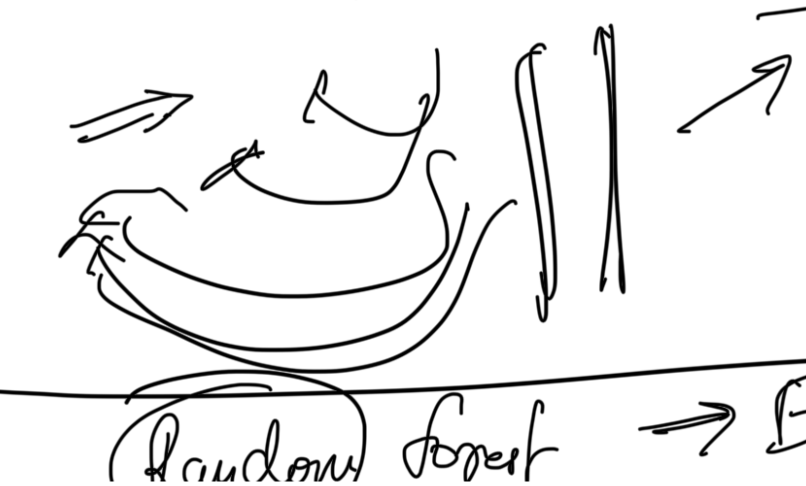
using some params.

Line? ⇒ m/c

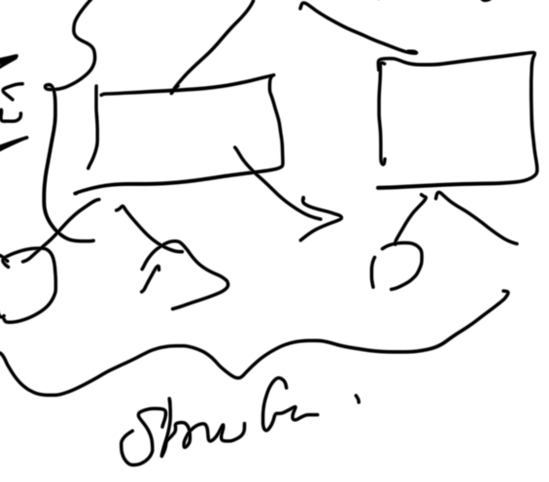
learns to make inference.

learning to make an via structure

→ learning to make inference



Structure

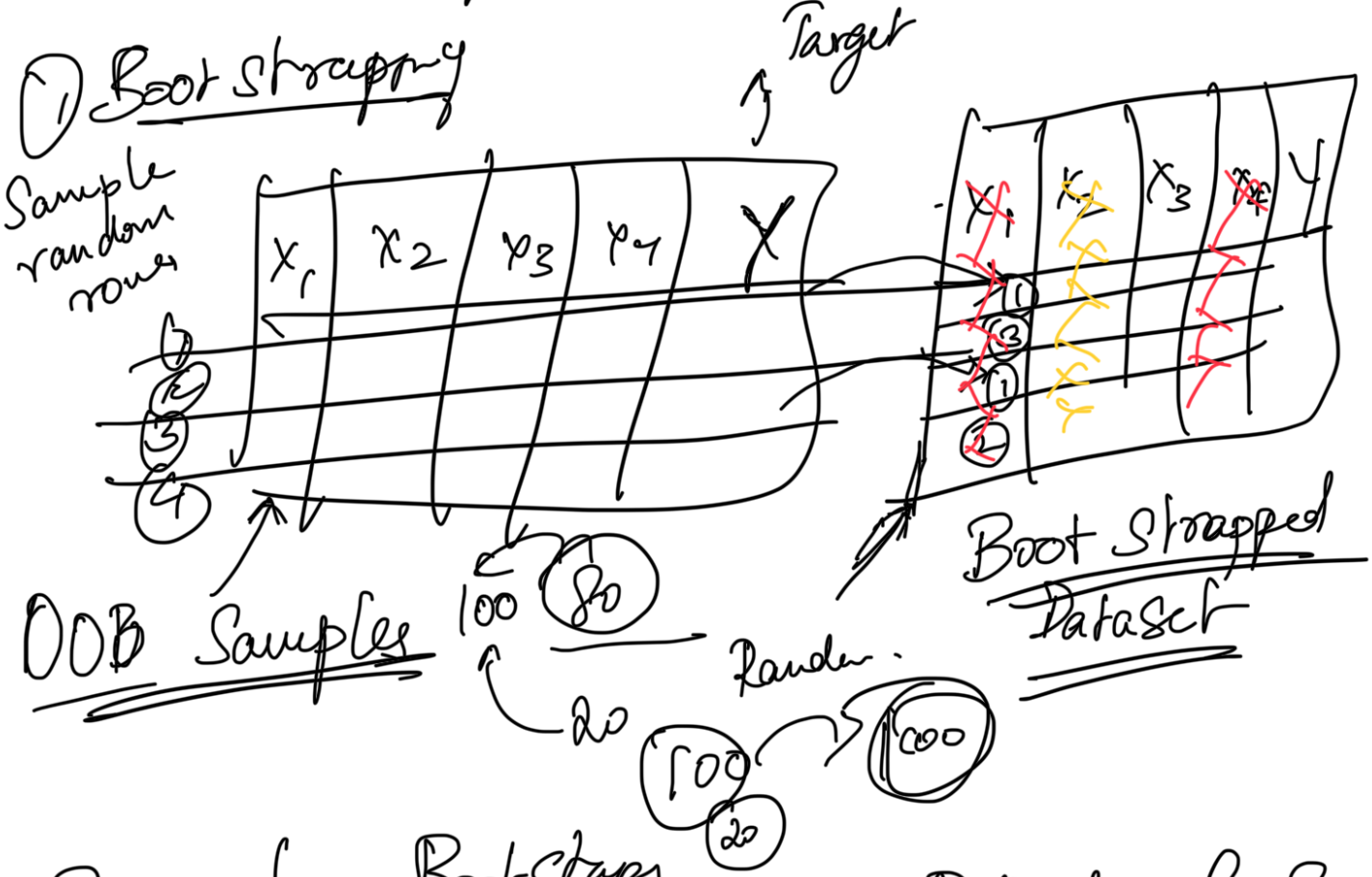


Structure.

Random Forest → Ensembling

Bagging

① Bootstrapping



Rules for Bootstraps
↳ Create a new Dataset of same size as original

② Forest
Create DT's using the Boot strapped Dataset

Tree ① Tree ② Tree ③ Tree ④

Data → Overfitting → Specific to rows & Colm

Random forest rows → random

Random set of Colm →

Tree 1 Tree 2 Tree 3 Tree 4 Tree 5 Tree n

x_1 x_2 x_3 x_4 y

NO

NO

Result

Yes	No
2	8

Aggregation

Bagging

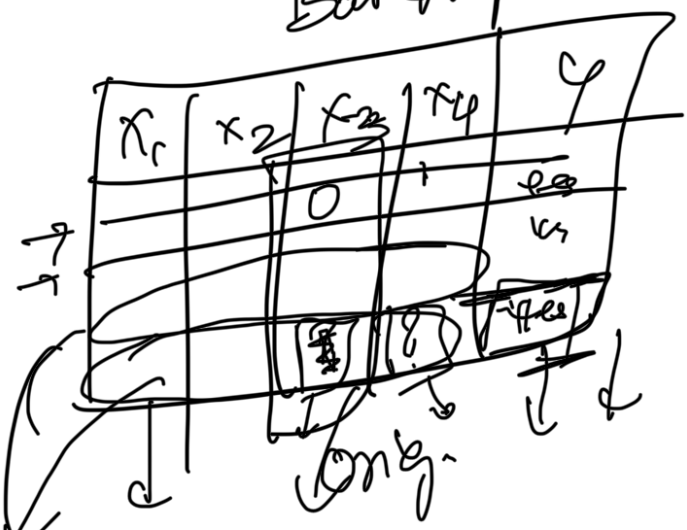
Bootstrap Aggregation

R.F ⇒ Extract Meaningful feats

OOB ⇒ Evaluation

Random Forest can also handle missing data

Bootstrap

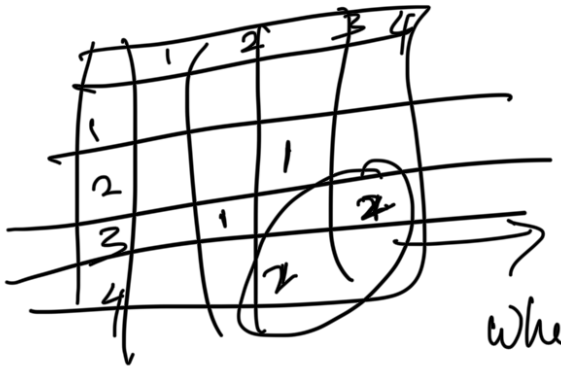


T_1 T_2 T_3 T_4 . . . . . .

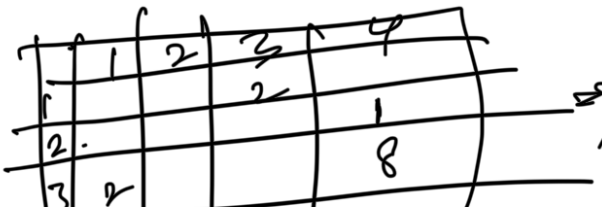
① Initial Guess (Mode / Median / Mean)

② Run it thru all the trees

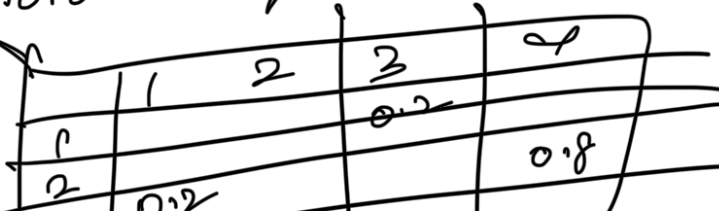
③ Proximity Matrix



keep adding when both the rows end in the same leaf



Total No. of same leaf



	1	8

3	0.1	0.8