

Recap: Decision Tree

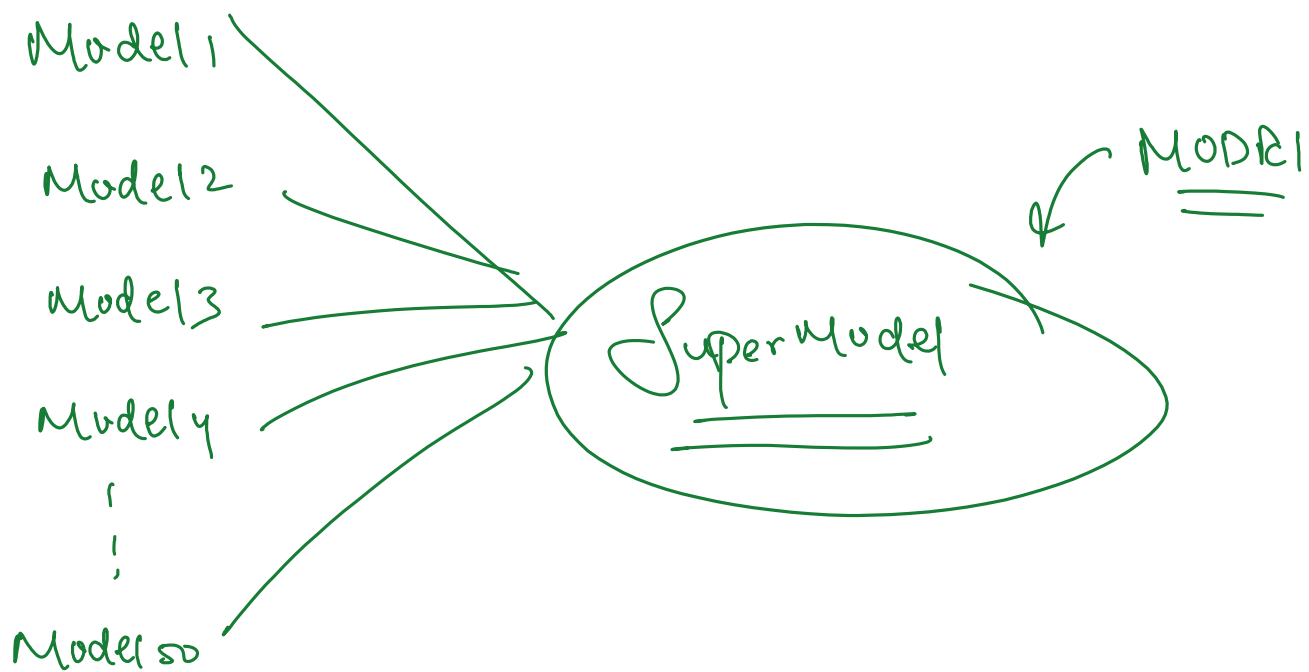
Practical implementation

→ Assignment → DTR.

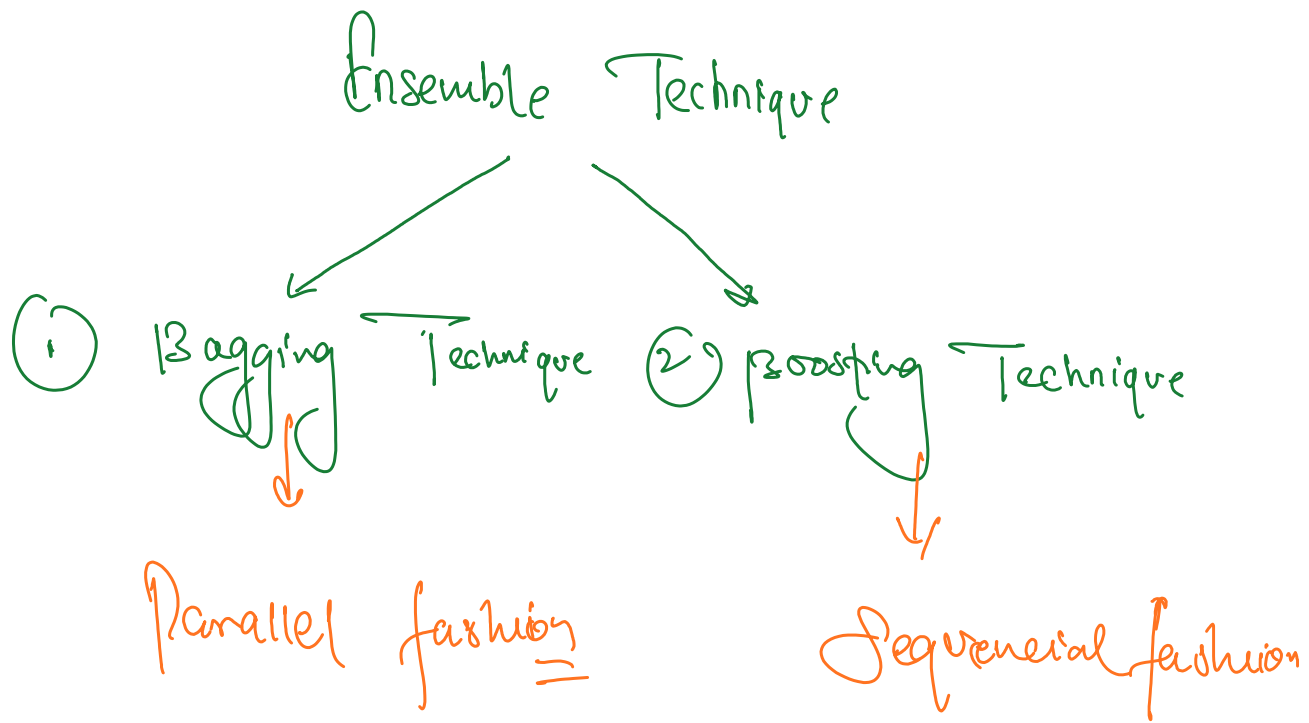
Agenda:

Ensemble learning

↓
Random forest ML Algorithm =



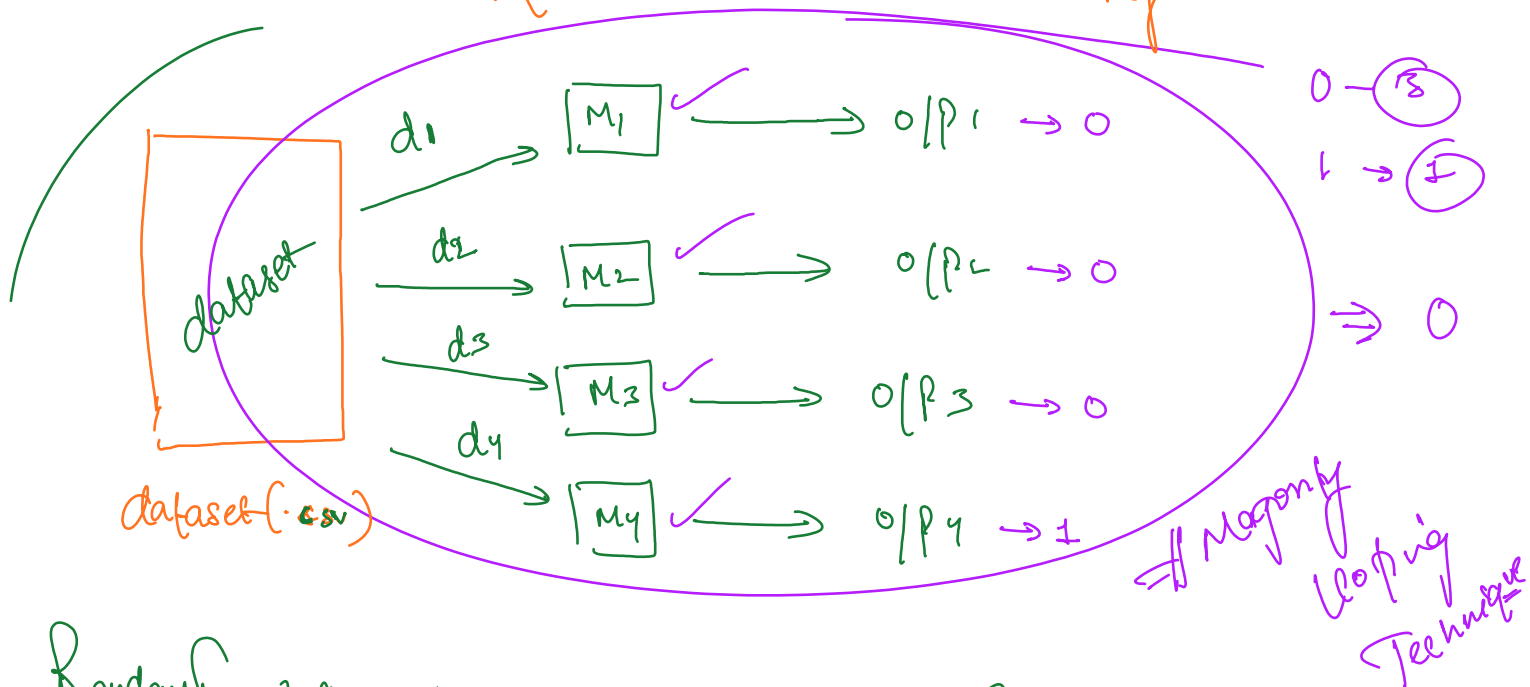
- * Ensemble method combine different- decision trees to deliver better predictive Result;
- * The primary principle behind the ensemble model is that a group of individual model gets combine to create a Master model (super model)



(1) Bagging Technique :

Here we will combine all the individual model parallelly.

The machine learning algorithm we use in Bagging Technique is "Random Forest ML Algo"



Random Forest : It is an Ensemble (Bagging Technique).

* In Random forest we combine multiple Ind. decision trees to create a master model.

Random Forest :



Regression

Clustering

Notes

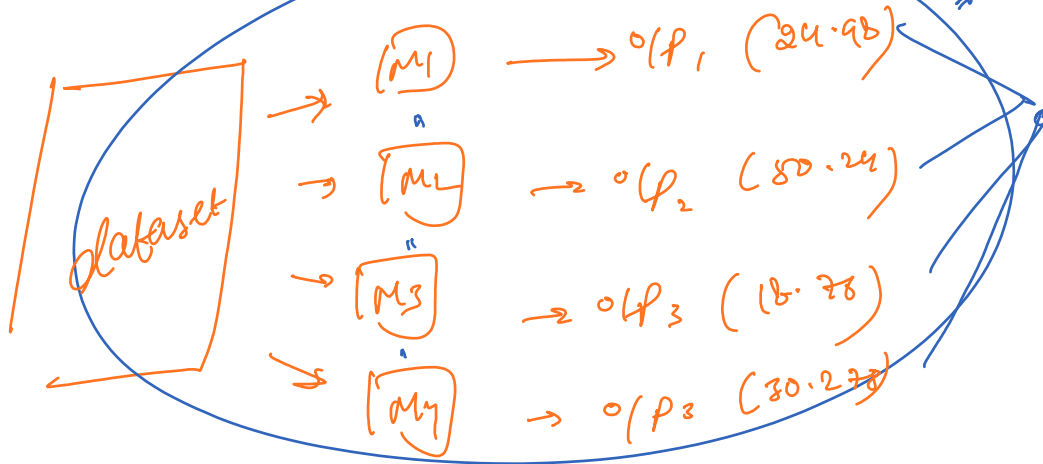
- Weak Learner/Model
- Individual Learner
- Base model
- Submodel
- Slave model

All same

- Master model
- Super model
- Combined model

All same

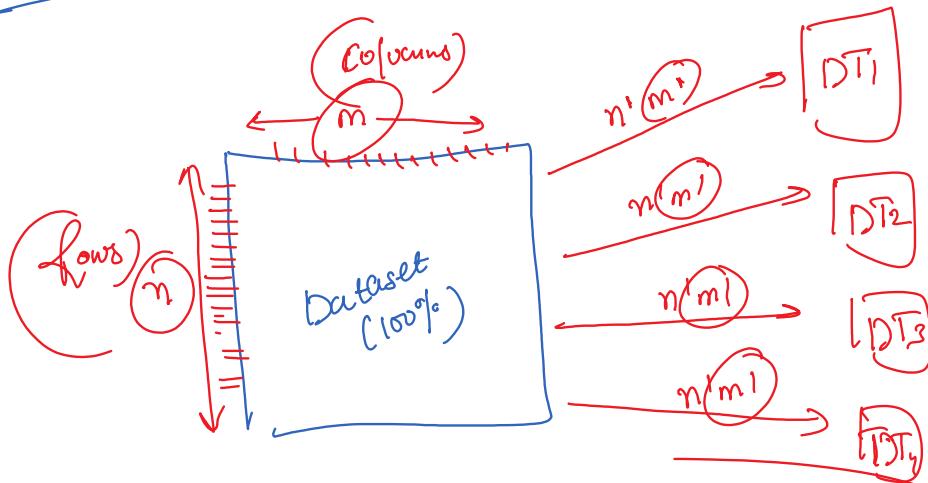
Regression \Rightarrow Continuous Output data



$$avg = \frac{24.98 + 80.24 + 18.78 + 30.278}{4}$$

Average
Technique

How dataset is divided for each individual model?



Here $n > n'$ and $m > m'$ True

- optimal size of Row and Column is \Rightarrow

(1) Generally we take $\frac{2}{3}$ of (Total Row)

(2) And $\sqrt{\text{Total column}}$

dataset

$(270, 16) \rightarrow$ complete data

↓

Weak Learner

Shape $(100, 4)$

$$\text{Row} = \frac{2}{8} \times \frac{90}{270} = 100 \leftarrow$$

$$\text{Covariance} = \sqrt{16} = \sqrt{4 \times 4} = 4$$