

Ensembles and its techniques

* Till now we have used only one ML Algorithm / model.

① * data → Model L → train - predict

② * data



Analogy

↳ D S Jobs → C (ML)

Ajay → A (ML)

(data Analytics) (ML)

→ One person might give you wrong advice.

→ You will connect to multiple mentors.

→ Chances of getting wrong is minimized.

* Ensembles : Combine multiple Models

: Prediction which is more stable and accurate as compared to individual Models.

Combine Multiple models

of Same Algorithm

- ▷ DT₁ (max depth: 5)
- ▷ DT₂ (max depth: 10)
- ▷ DT₃ (max depth: 12)

of different algorithms

- ▷ Logistic Regression
- ▷ SVC
- ▷ DTC

* Ensemble :- Not necessarily only one type of algorithm.

Ensemble technique

Parallel
Technique

(Bagging)

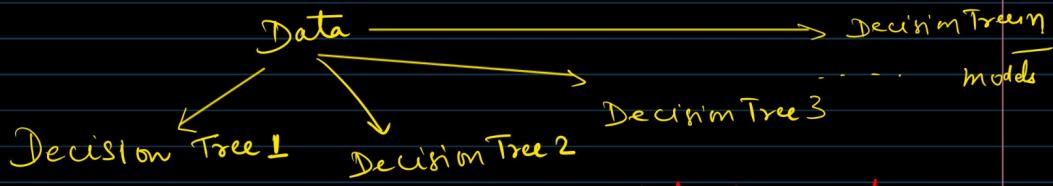
Sequential
Technique

(Boosting)

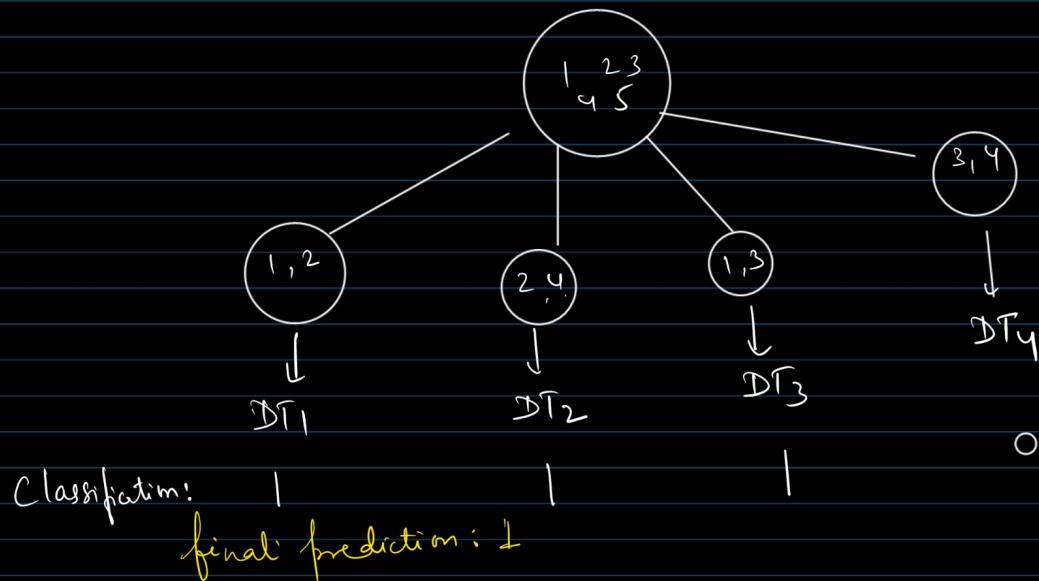
parallel + sequential
Technique

(Stacking)

Parallel technique of Ensembles



→ All the models are built **parallelly** and **independent** of each other.



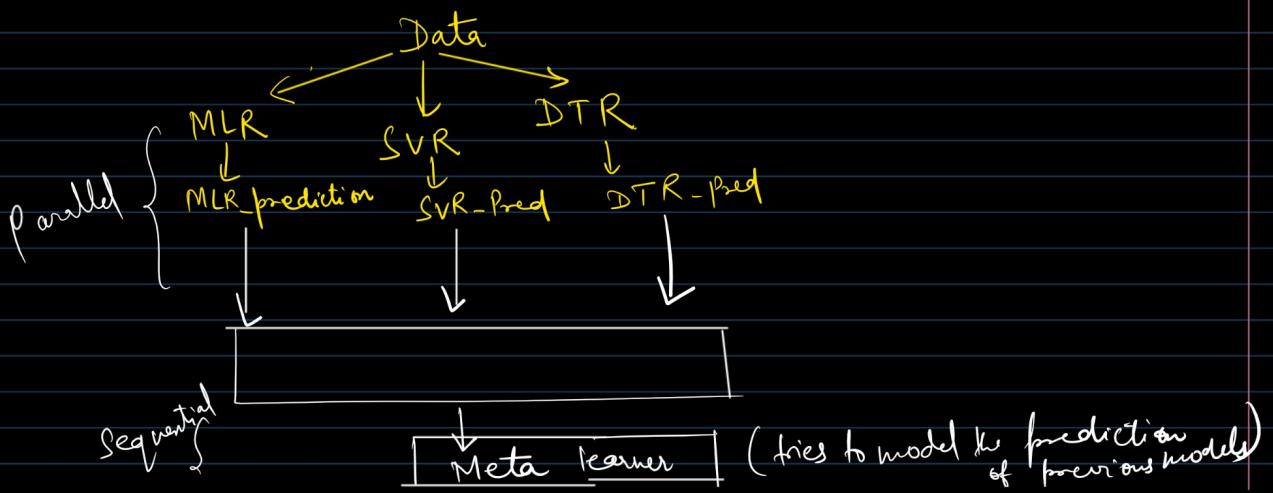
* Sequential technique of Ensembles

→ All the models are built **sequentially** and **dependent** on each other.

→ learning from mistakes of previous model.



* Parallel + Sequential technique of Ensembles



L
↓
Predict the final output

Ensemble Technique

Parallel

Sequential

Parallel + Sequential

Ensemble Models

Bagging
(parallel technique)

Boosting
(sequential technique)

Stacking
(both technique)

Multiple type
of algorithms

Only one type of
Algorithm is used

* Bagging :> Random forest $\begin{cases} \text{Regressor} \\ \text{Classifier} \end{cases}$

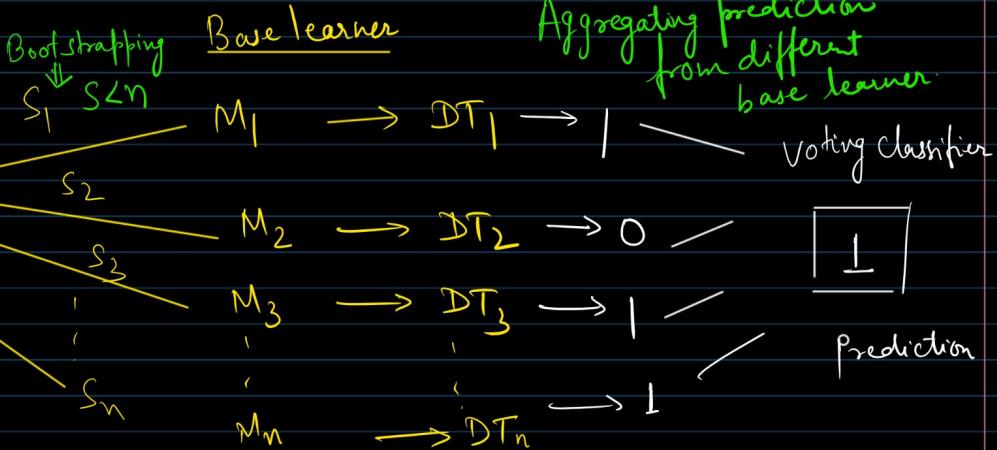
Custom Bagging (different Algorithms)

* Boosting :- ① AdaBoost
② Gradient Boost
③ XG Boost
④ Cat Boost.

* Stacking :- We will see example of different models.

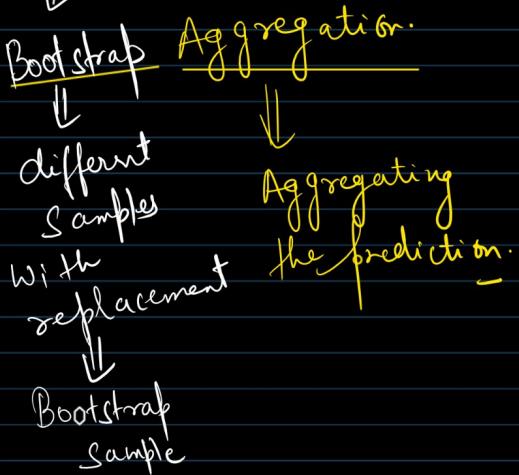
* Bagging technique (with classification problem)

→ Parallel Models

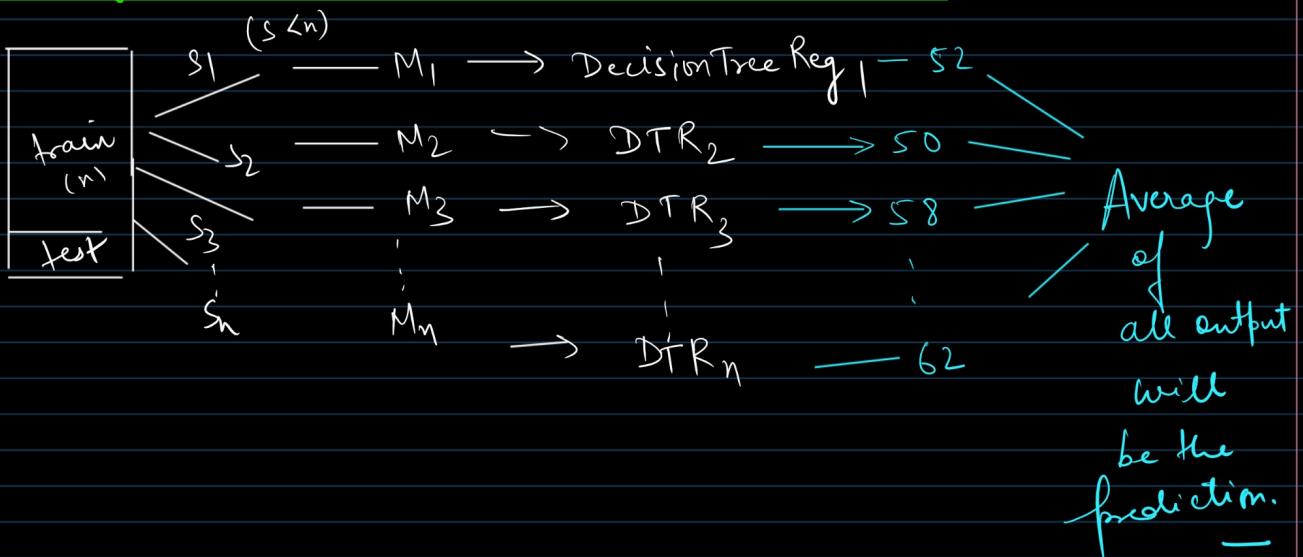


* Samples (subset) for each model is taken with replacement.

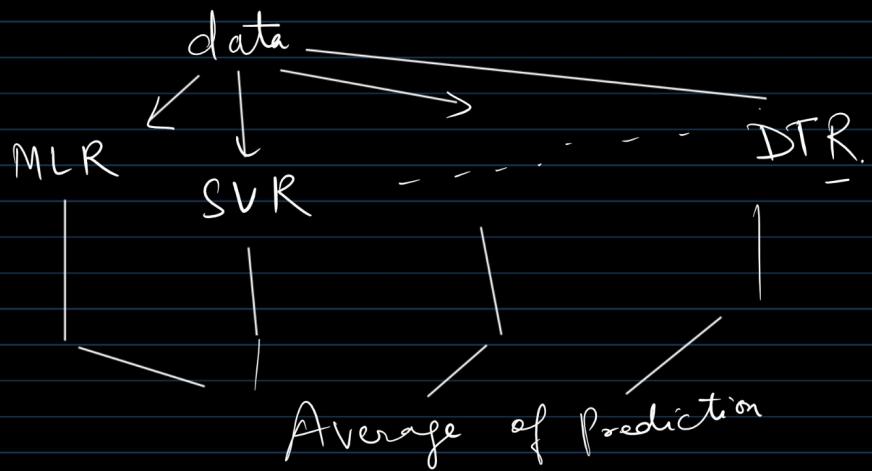
Bagging



* Bagging technique with Regression problem.



* Custom Pseudo Bagging technique (Building multiple models)



* Random forest (Classifier, Regressor)

* Custom Bagging (Multiple Regression & Classification model).

(Also pipeline | column Transformer | Missing Value, Encoding, scaling)