

Agenda

- Decision Trees
- Random Forest
- Bagging
- Boosting

CART (Classification and Regression Trees)

Regression Trees

Features: any features
Label: continuous

Error function:

① MSE → Mean Square Error

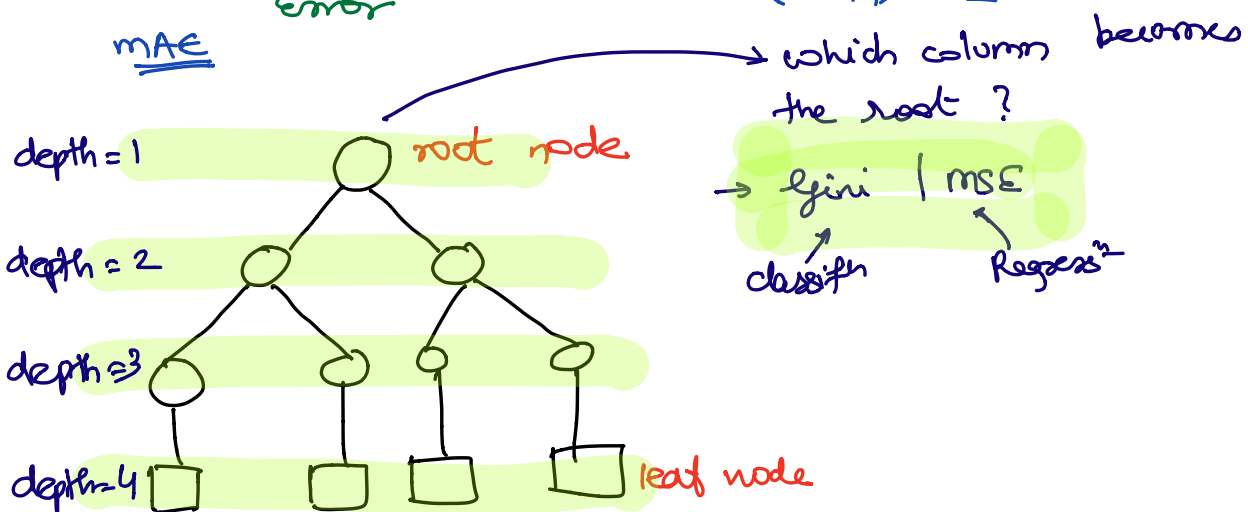
MAE

Classification Trees

Features: any features
Label: categorical

Error function:

① Gini impurity.
(Entropy) → IG

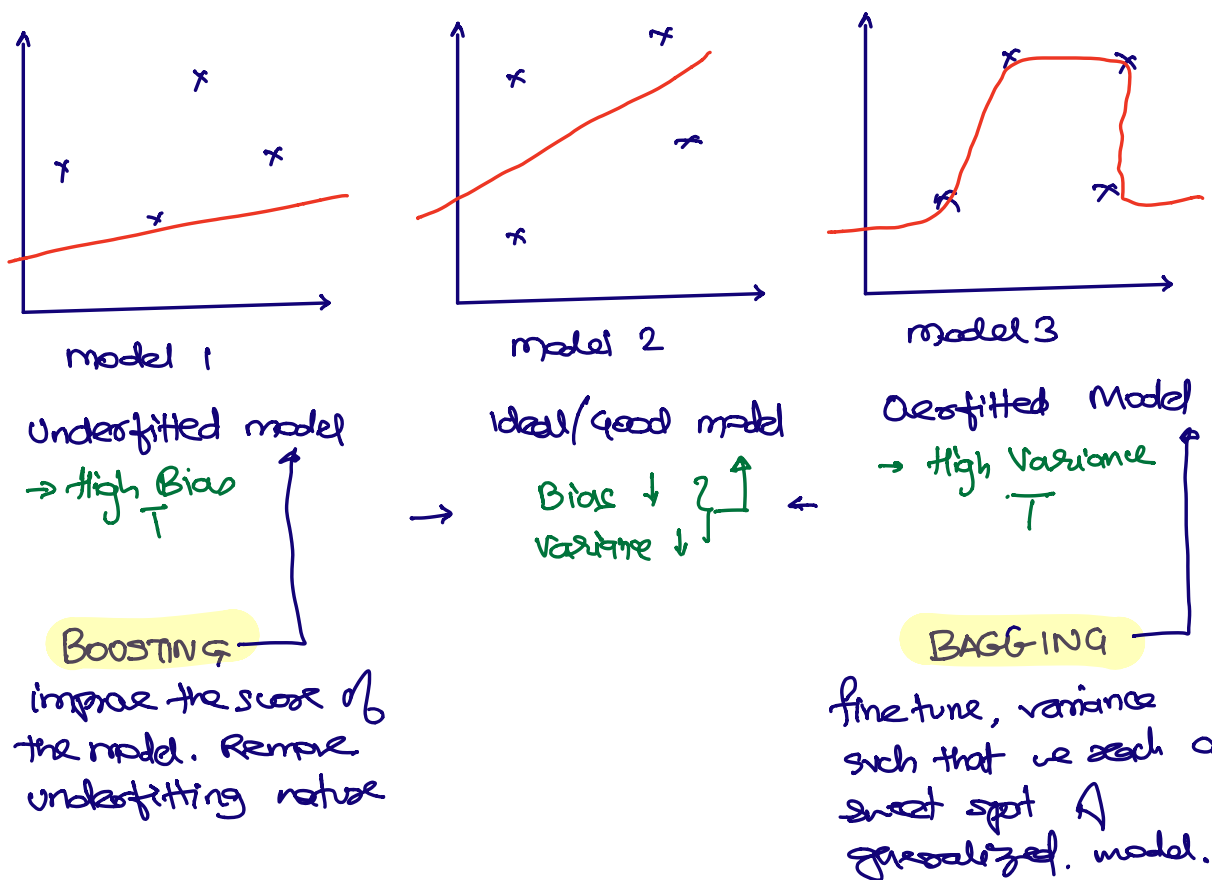


ML Engineer:

hit & try to find sweet spot

- ① Ensure you control the max depth else the algorithm will memorize the data.
- ② Don't use Decision Tree for large dataset.

Bagging and Boosting framework.



Concept of Bagging

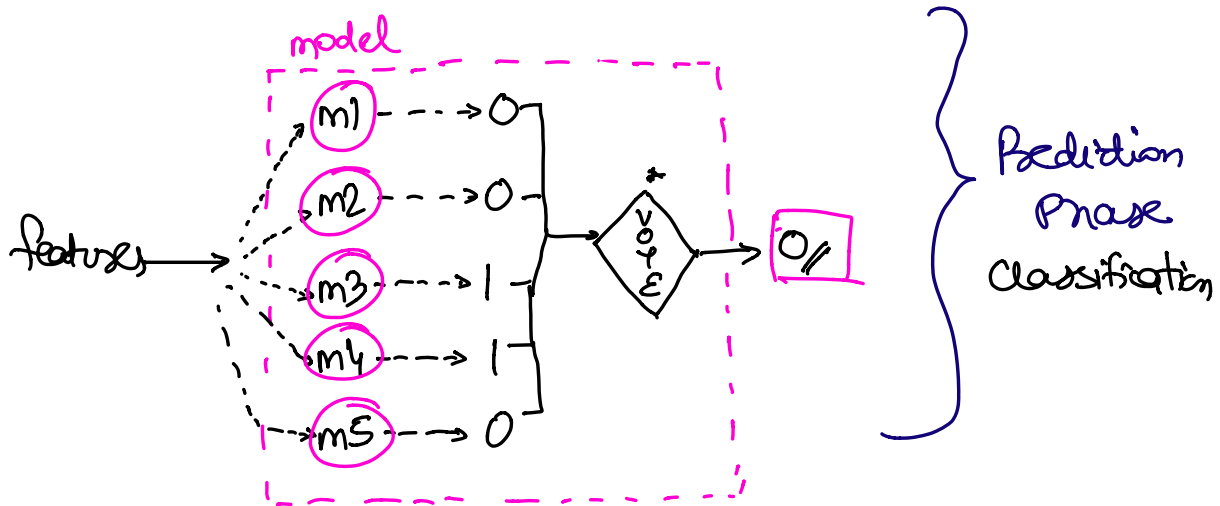
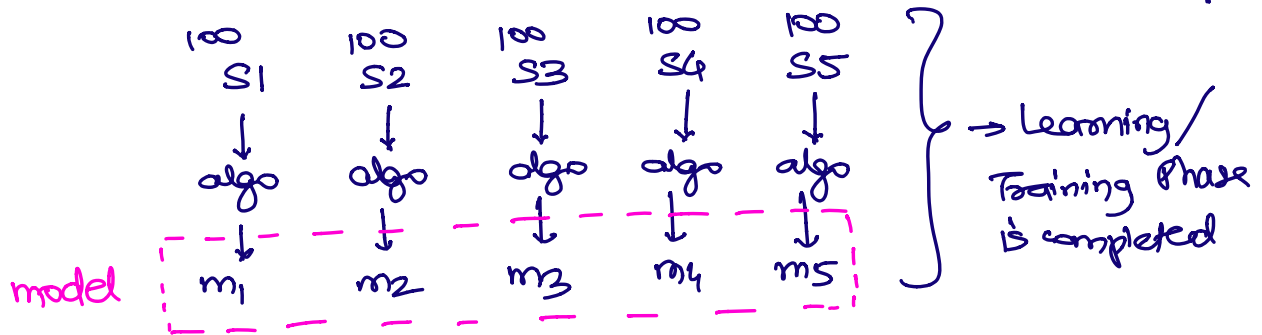
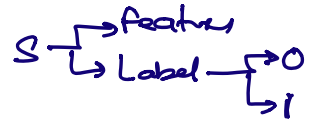


Bagging \rightarrow Bootstrap Aggregation

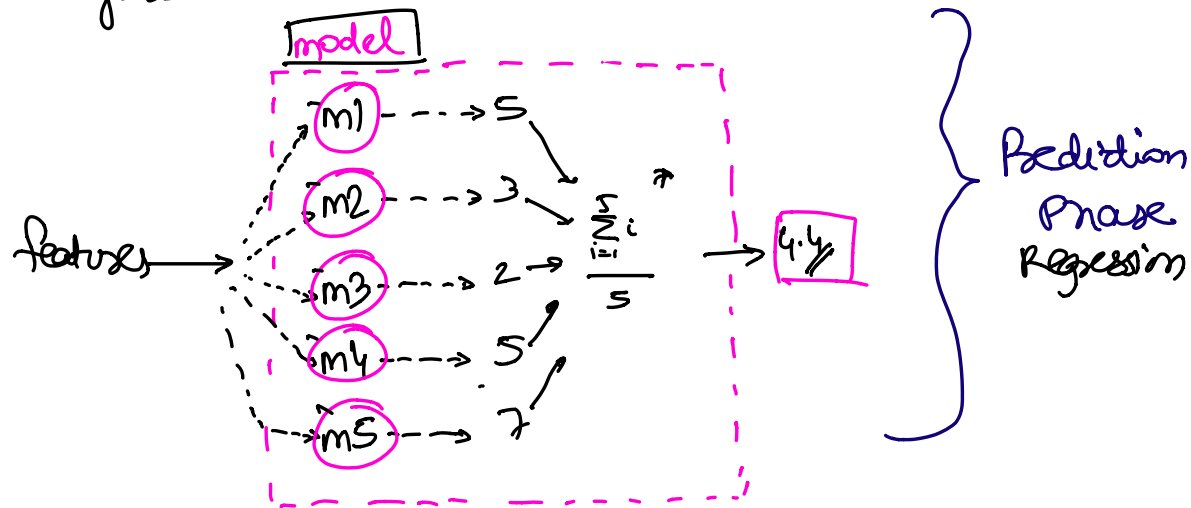
no_estimators = 5 ... odd no. of estimators (binary class)
dataset → 100 rows
S → Feature

algo. = Decision Tree Classifier

dataset \rightarrow 100 records



Regression Scenario

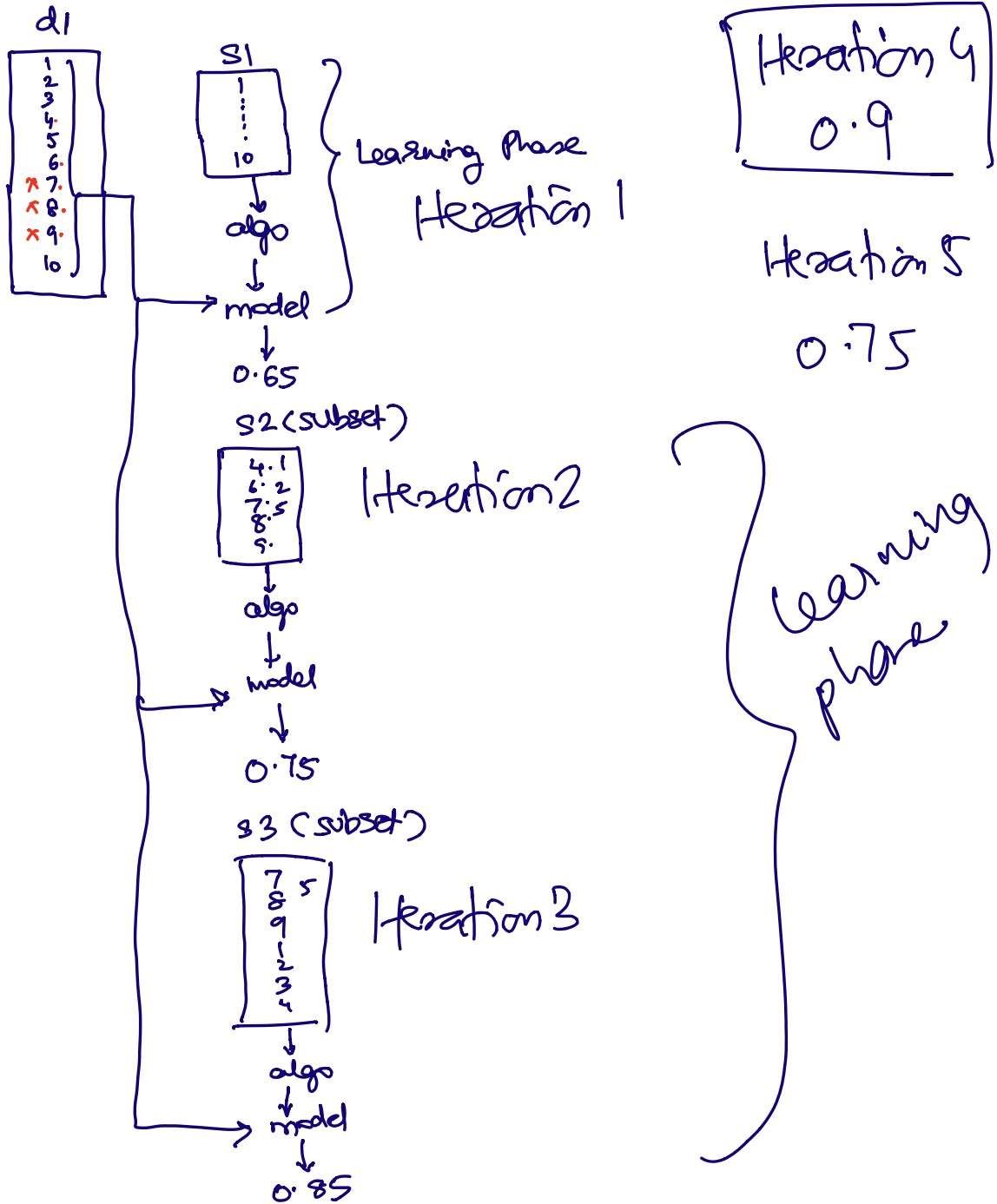


Boosting : (Note: used mostly in Neural networks)

- applicable to improve an under-fitted model.
- helps improving the score of the model.
- may lead to Overfitting.
- Ideally, Boosting is used only for classification usecases though it do support Regression.

no_estimators : 5 (\leftarrow 5 iterations)
algo: K-NN (test_size : 0.2)

dataset



model (0.9)
↑
finalized model by
Boosting

Random forest

↳ Decision tree with Bagging.

- set of decision trees participate in prediction phase.
- classification → voting
- regression → mean