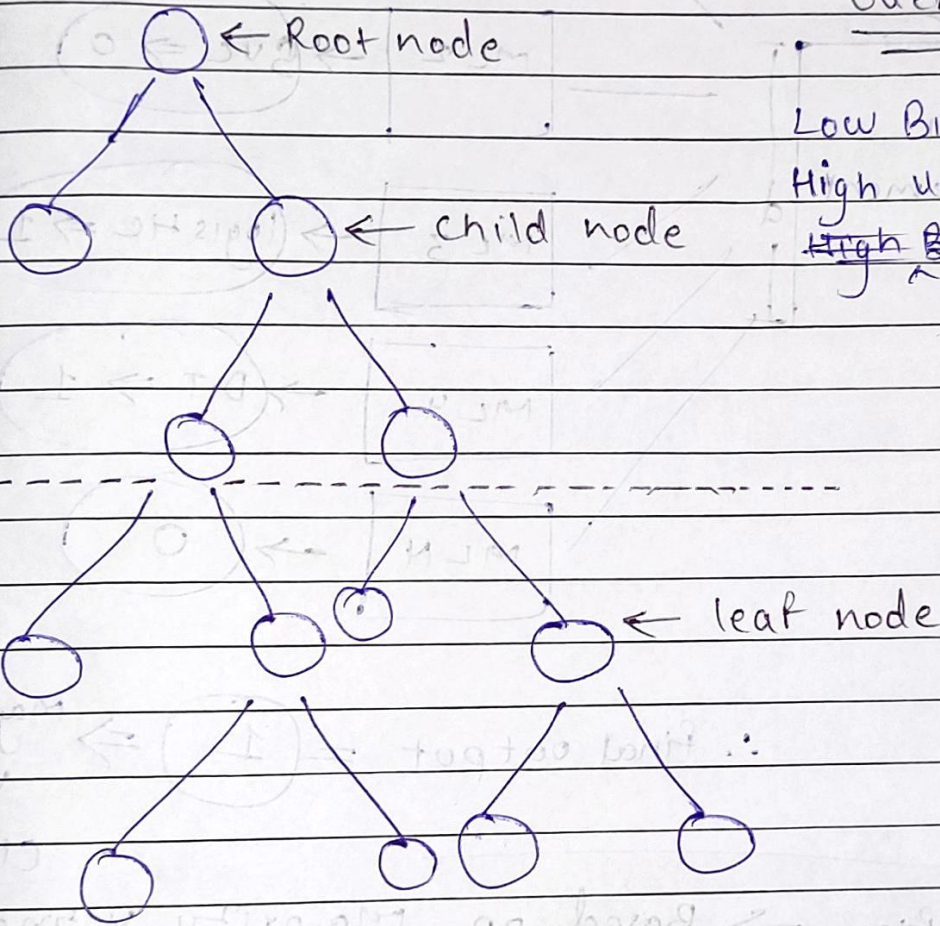① Random forest classifier
② Random forest Regressor

Decision Tree :-

overfitting

prepruning

Low Bias
High variance ⎤ Post pru-
~~High Bias variance~~ ⎦ ning
High Bias ~~variance~~

← Root node

← child node

← leaf node

overfitting

Low Bias ⎤ Pre pruning
        ⎥ ⇒
High variance ⎦ Post pruning

* Bagging and Boosting :-

(1) Bagging :- [Ensemble Techniqu

(Bootstrap
Aggregation)

Bagging of Algorithm

ML 1 → D.T ⇒ 1

ML 2 → SVC ⇒ 0

ML 3 → logistic ⇒ 1

ML 4 → D.T ⇒ 1

MLN → 0

Problem statement

d

∴ final output = 1 ⇒ majority voting Classifier

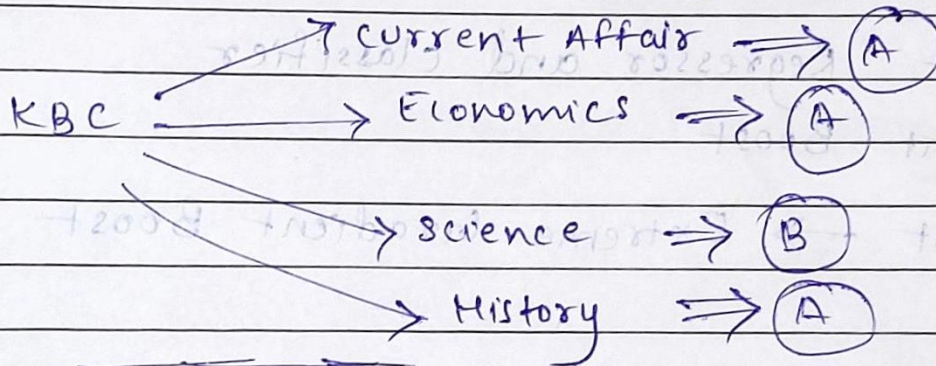* In classifier ⇒ Based on majority voting

In Regression ⇒ Based on mean/Average

Example :-

KBC [Amitabh Bachan]

{
Krish ⟶ DATA SCINCE

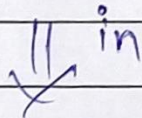UPSC ⟶ Diversified ⟹ Multiple Domain
}

UPSC person can answer easily in KBC.

KBC ⟶
- current Affair ⟹ (A)
- Economics ⟹ (A)
- science ⟹ (B)
- History ⟹ (A)

| Essemble Techniques |

‖ in

(Kaggle, Hacker Rank)

⇓

(overperforming result)

| final ⟹ A |

↑

[majority voting]

for Regression we use mean instead of majority.

② Boosting :-

weak learners
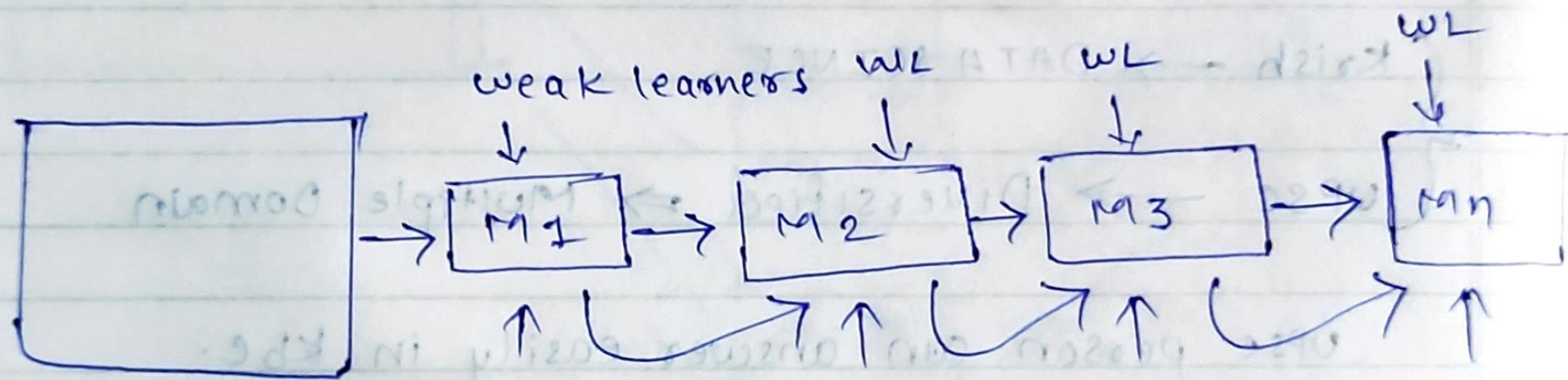
① AdaBoost Regressor and classifier
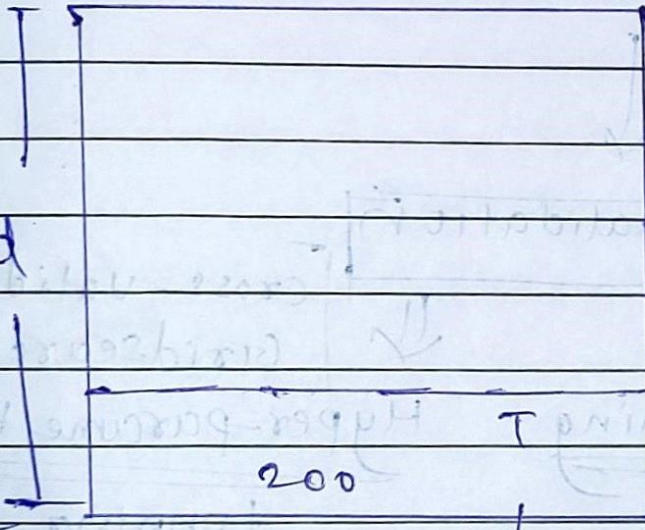
② Gradient Boost

③ xgboost → Extreme Gradient Boost

Random forest classification and Regression:-

with replacement

$d_1 < d$

$2m_1 < m$

$\vdash$ — m — $\dashv$

ML 1 → DT1

d1

Raw sampling + feature sampling
206

ML 2 → DT2

Raw sampling + feature sampling
200

1000 | d

Raw sampling + feature
200

ML 3 → DT3

200

200

200

→ validation
data

ML N → DTN

validating

oob-score = True
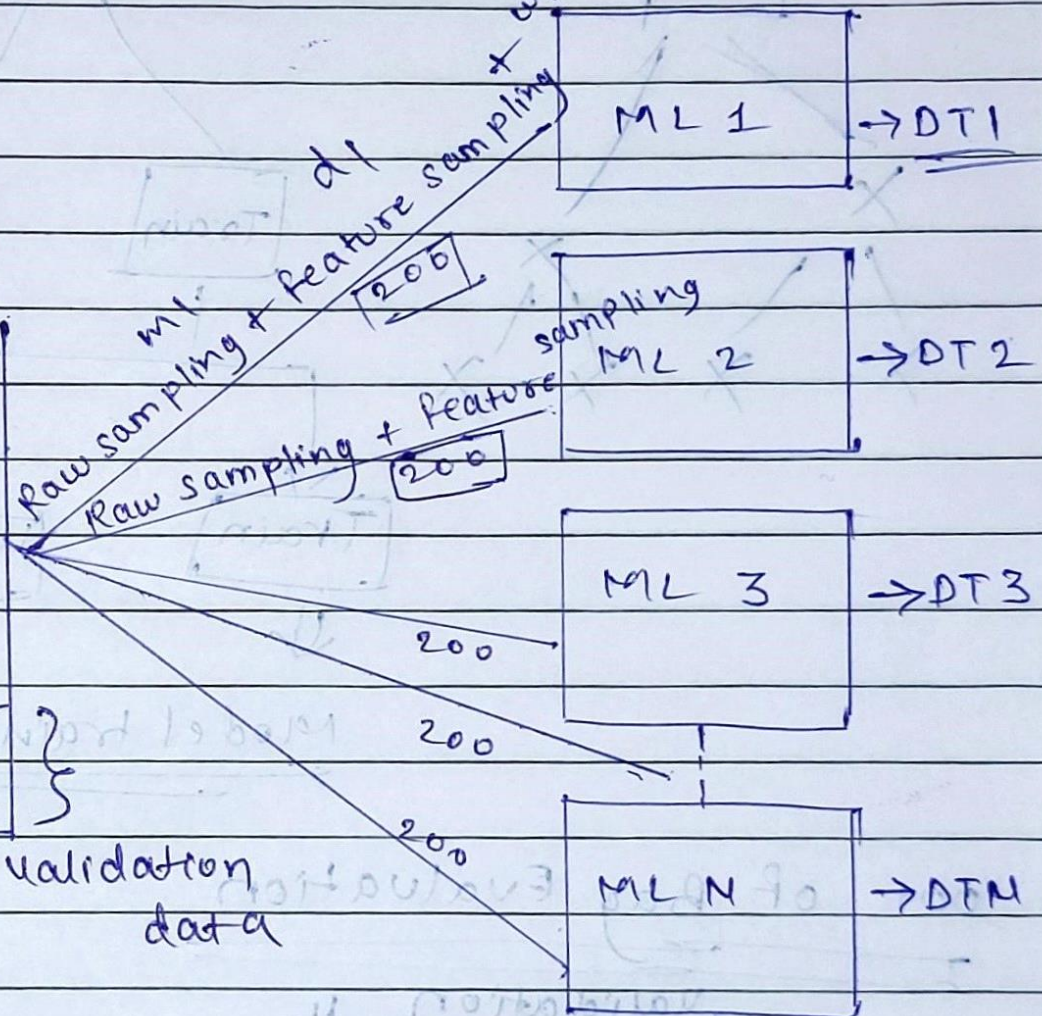
DT₁ ⇒ overfitting ⇒ Pre pruning or post pruning
⇒ cubersome task

① Low Bias ⟹ Training

② High variance ⟹ low variance

DATASET

OOB-Score = True

Train   Test m ⟹ , m sklearn

$p > i$ p

m —

Train   Validation

cross-validation
Gridsearch CV

Model training   Hyper-parameter

tunning

Out of Bag Evaluation

Validation

oob-score = 0.83%

score

out of Bag Error

1 - oob-score = 1 - 0.83 = 0.17% } validation Error