Ensemble Technique.

1. Bagging ⌣
2. Boosting ⌣
3. Stacking
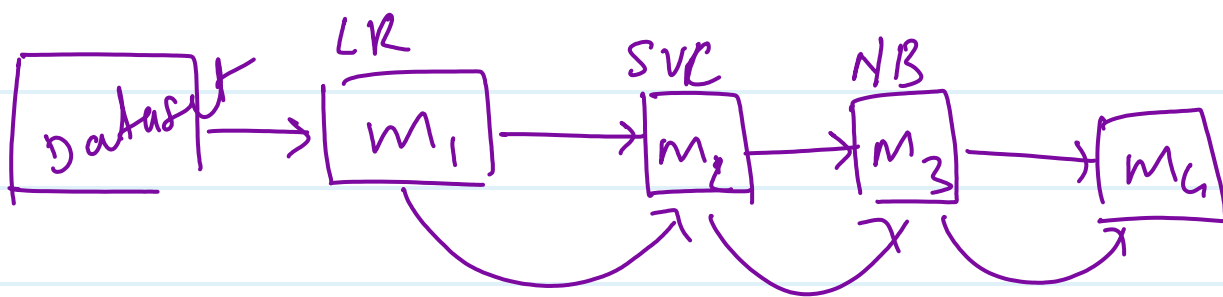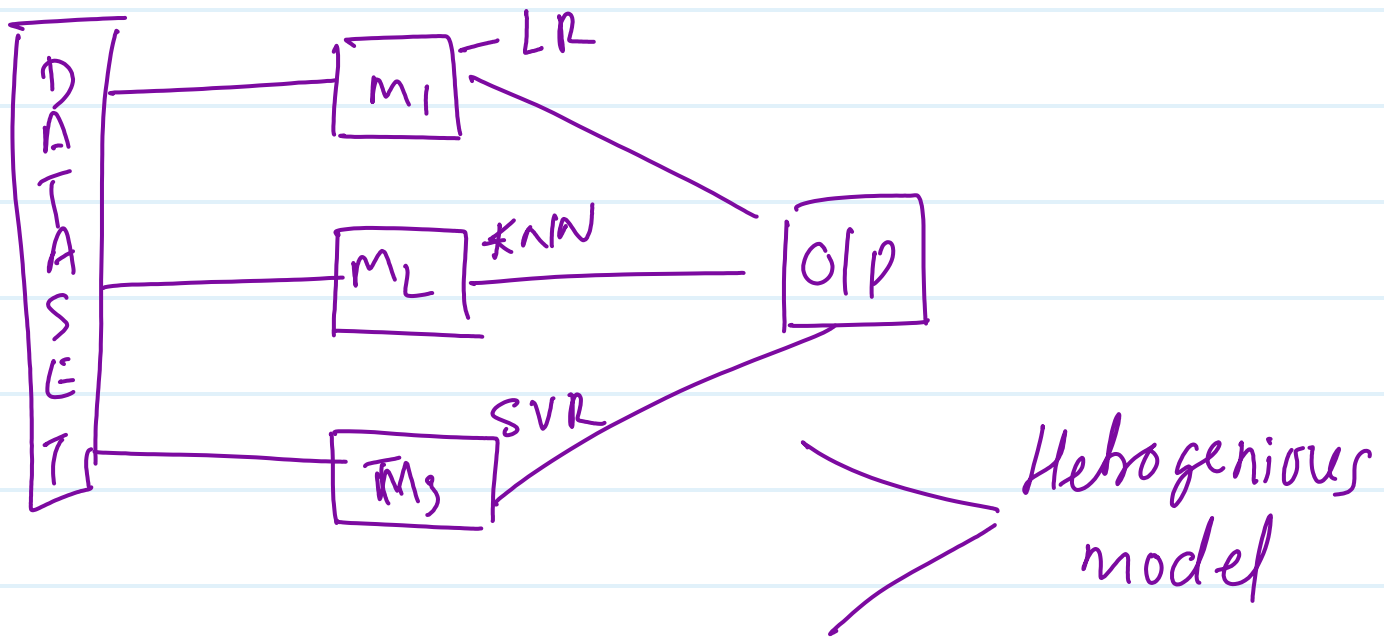
DATASET → M_1 (LR), M_2 (*NN), M_3 (SVR) → O/P

Heterogenious model

Dataset → M_1 (LR) → M_2 (SVC) → M_3 (NB) → M_4

Bagging

① Random forest

Dataset

X_1	X_2	X_3	X_4	X_5	X_n	Y

Row — 1000
Coloum — 6

D

$R_{200} + C_3$ → $\boxed{M_1}$ DT $0/2.1$

$R_{200} + C_3$ → $\boxed{M_2}$ DT $0/3.6$

$R_{200} + C_3$ → $\boxed{M_3}$ DT $1/3.2$

d' → $\boxed{M_n}$ DT $0/4.8$

classification
vote based
O/P O

Regression
Avg. value

$$\frac{2.1 + 3.6 + 3.2 + 4.8}{4}$$

$\Rightarrow 3.\text{---}$

$D > d'$

↑
Base learners

$DT \begin{cases} \text{low baised} \\ \text{high variance} \end{cases}$

$RF \begin{cases} \text{low baised} \\ \text{low variance} \end{cases}$

in bagging model train porellly

� Boosting

① Ada boost
② Gradient boost
③ XG boost (Xtream Gradient boost)

correct predicted point

```
1000              400              350            200         50
┌──────┐        ┌──────┐        ┌──────┐      ┌──────┐    ┌────┐
│ Data │  ───▶  │  M₁  │  ───▶  │  M₂  │ ───▶ │  M₃  │───▶│ M₄ │
└──────┘        └──────┘        └──────┘      └──────┘    └────┘
```

weak learner

600
wrong predicted point
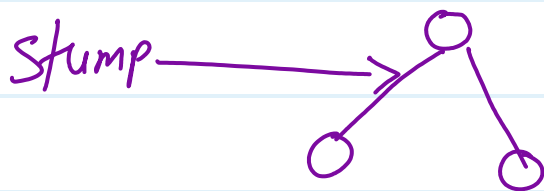
250   50
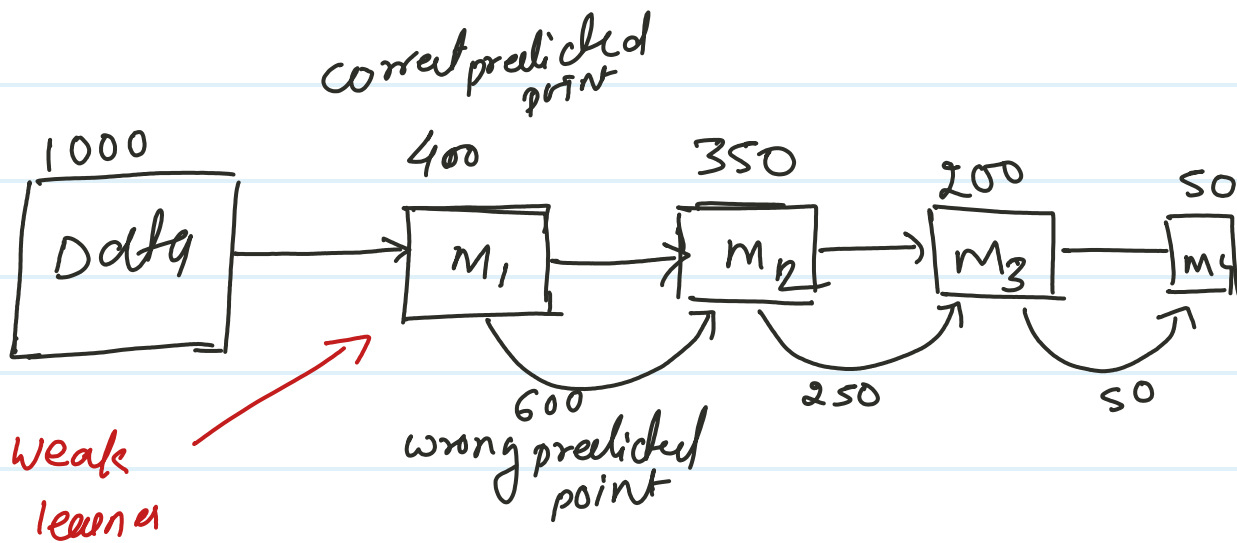
"Sequence training"

$$F = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \alpha_4(M_n)$$
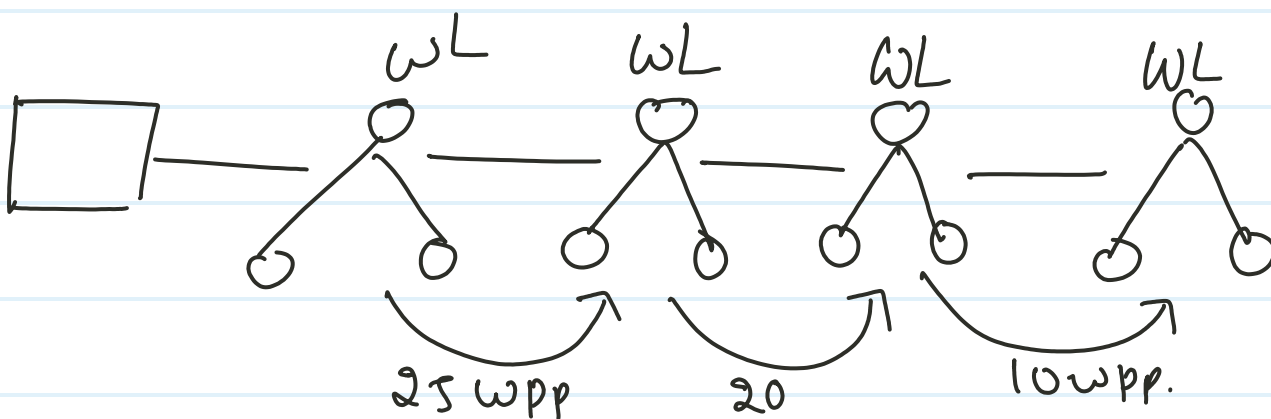
$\alpha =$ weight

① ADA Boost :-

→ most of time used to build classification problem.

→ It only use DT algorithms for model building

→ weaklearner add weight at the o/p of the WL.
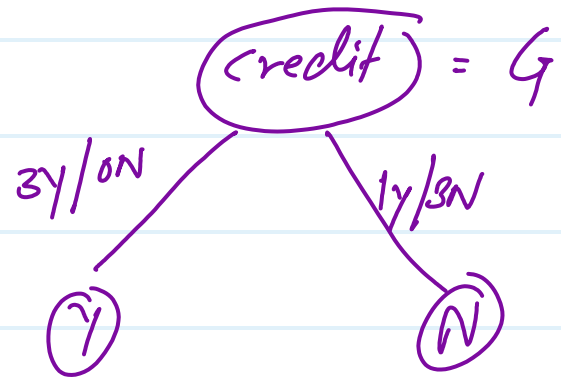
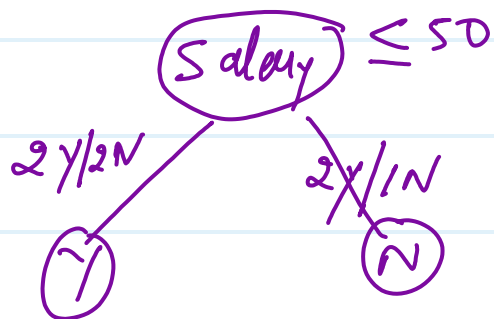Stump ⟶

DT with on depth only called stump

WL   WL   WL   WL

25 WPP   20   10WPP.

$$F = \alpha_1 (m_1) + \alpha_2 (m_2) + \alpha_3 (m_3) + \cdots \alpha_n (m_n)$$

$$\alpha = weight.$$

salary	credit	Approach	weight assign
<=50k	B	N	1/7
<=50k	G	Y	1/7
<=50k	G	Y	1/7
>50k	B	N	1/7
>50k	G	Y	1/7
>50k	N	Y	1/7
<=50k	N	N	1/7

step-I   we created DT stump

we calculate either entropy or Gini on based it we calculate information gain for root node.

step ②   (Credit) = G

3Y/0N   1Y/3N

Ⓨ   Ⓝ
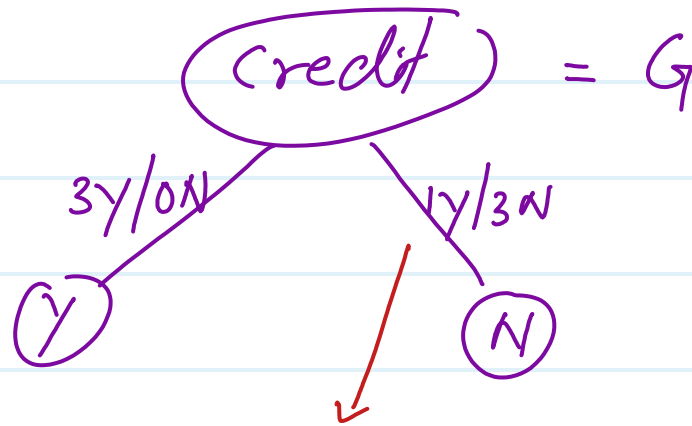
in that datapoint I is wrongly predicted point.
So we find α (weight)

$$formula = \frac{1}{2}\log\left[\frac{1-T\epsilon}{T\epsilon}\right]$$

step - ③   one wrongly predicted

$$performance\ of\ stump = \frac{1}{2}\ln\left[\frac{1-T.\epsilon}{T.\epsilon}\right]$$

$$= \frac{1}{2}\ln\left[\frac{1-\frac{1}{7}}{1/7}\right]$$

$$\left[sum\ of\ total\ error = \frac{1}{7}\right]$$

$$\approx 0.896$$

$$\alpha_1 = 0.896$$

Ada boost

① Regression

② classification

mainly is used to make binary class classification problem.

Alway use DT as a weak learner.

DT will alway used as one depth called DT stump.
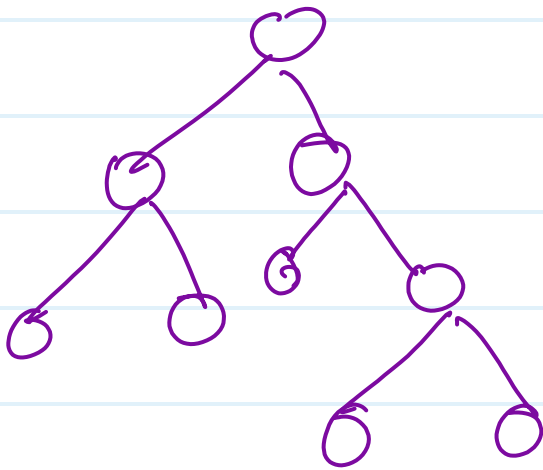
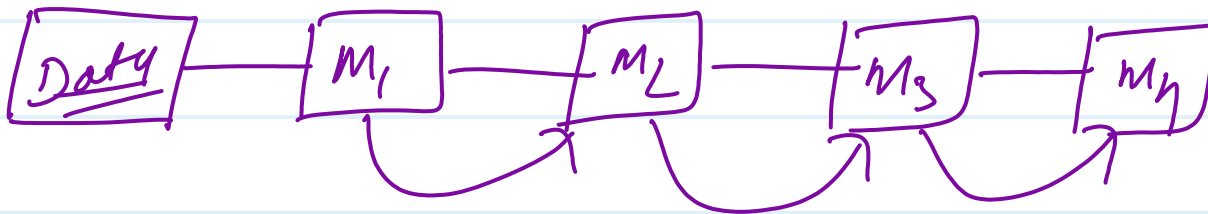# Gradient boost and xtream Gradient boost

① Regression
① classification

mainly it used to build classification

problem, binary and multiclass classification.

✳ DT as a weak learner.

It create complete decision treen.

$$xG\,boost = \alpha_1 M_1 + \alpha_2 M_2 + \cdots \alpha_n m_n$$