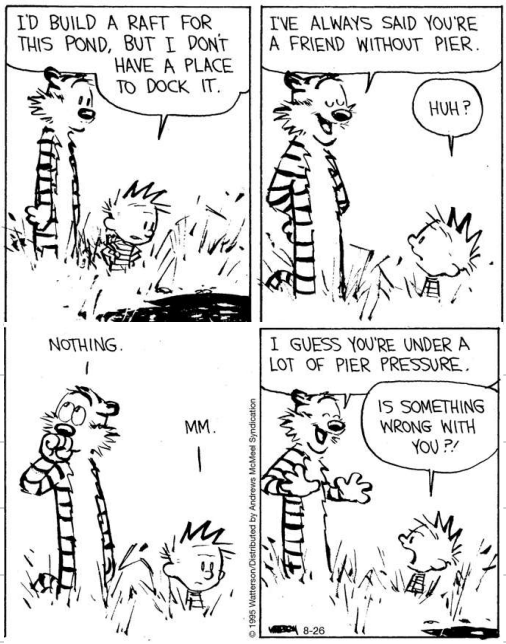
Session - 6

Boosting - 1

Sep 05, 2025



Agenda

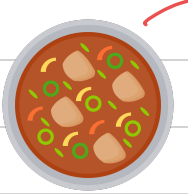
1. Additive Combination
2. Intuition of Boosting
3. Steps involved
4. Pseudo Residuals
5. Algorithm.

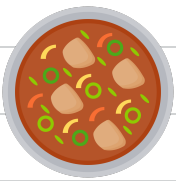
Bagging & Boosting

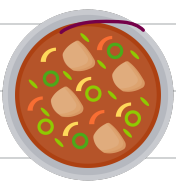
Bagging : Base Learners (overfit) + R.S. + C.S (Randomly.) + Agg. (majority voting.)

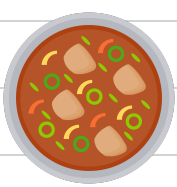
Boosting : Base Learners (underfitting) + Additive Combination

MasterChef - Androids!!

1:  → Forgot to add Power

2:  → Forgot to cut Power

3.  → Added too much salt

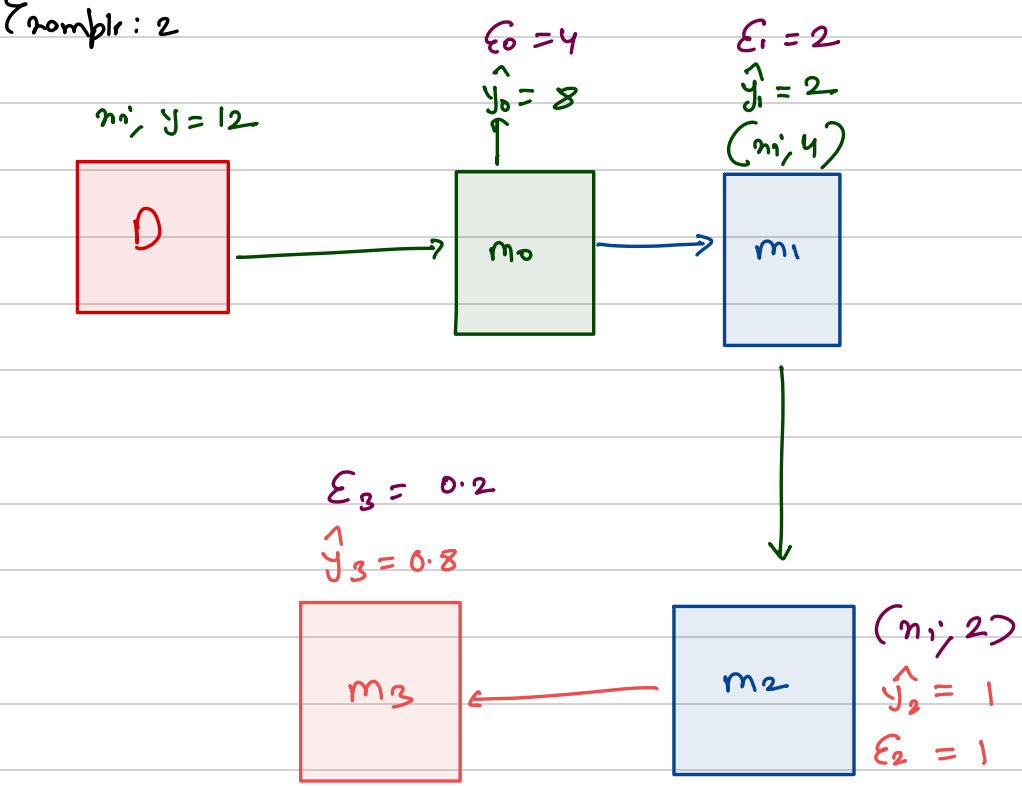
4  → Quite decent

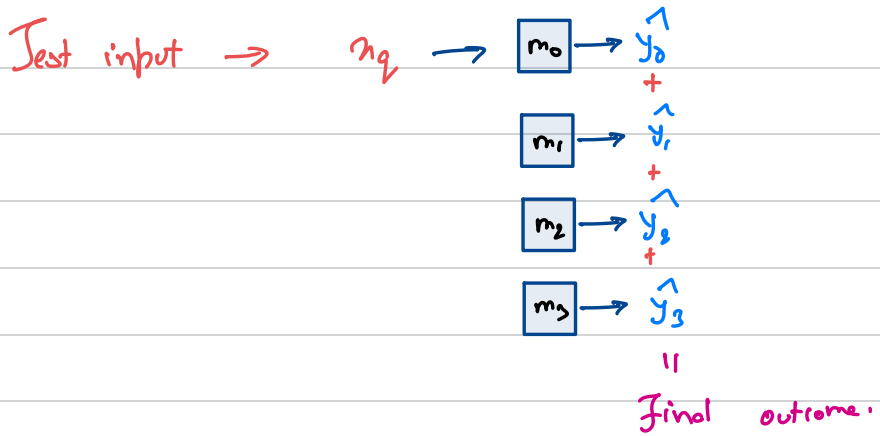
model # 0 — Error made by model 0 — DT with depth 1 — (Age, ε_0) — (Age, ε_1) — (Age, ε_2)

Boosting

Age (x_1)	Price (y_i)	m_0	ε_0	m_1	ε_1	m_2	ε_2	m_3
5	17	10.6	6.4	5.5	0.9	0.8	0.1	
2	20	10.6	9.4	8.0	1.4	1.1	0.3	
10	8	10.6	-2.6	-1.5	-1.1	-0.8	-0.3	
15	5	10.6	-5.6	-1.1	-1.1	-0.9	-0.2	
20	3	10.6	-7.6	-1.6	-1.6	-1.1	-0.5	

Example: 2





Which of the following model are underfit models ?

0 users have participated

✓	A	Decision Stump (Decision Tree with depth = 1)	0%
	B	Decision Tree with depth = 20	0%
	C	Decision Tree with depth = 10	0%
	D	None of the above.	0%

End Quiz Now

Leaderboard

Based on all quizzes from the session

2	Souvik Adhikary	1/1	91.53
1	Shoreya gupta	1/1	94.86
3	Sri Harsha Nanduri	1/1	89.87
4	Sumanth Andhavarapu	1/1	88.47
5	Samyucktha Ramesh	1/1	88.26
6	Mohanakrishna	1/1	88.20
7	Snehal Adhikary	1/1	87.73
8	Kiran Hebasur	1/1	87.60
9	Rakesh Karade	1/1	86.66
10	Nitin	1/1	86.47

What model will have high bias low variance ? (Hint: think of simplest model)

0 users have participated

✓	A	Mean model (which predict mean value everytime)	0%
	B	Random model (model which predicts random value everytime)	0%
	C	A DT fit perfectly on training data	0%

End Quiz Now

2	Souvik Adhikary	2/2	182.60
1	Shoreya gupta	2/2	184.93
3	Snehal Adhikary	2/2	173.00
4	Sumanth Andhavarapu	2/2	172.20
5	Perisetla Pavan Kalyan	2/2	170.46
6	Kiran Hebasur	2/2	169.07
7	Samyucktha Ramesh	2/2	167.00
8	SHASHANK JHA	2/2	162.60
9	Sri Harsha Nanduri	2/2	162.00
10	Karthik	2/2	159.74

What will be the training error for high bias model ?

0 users have participated

✓	A	High error	0%
	B	No error	0%
	C	Low error	0%

End Quiz Now

2	Shoreya gupta	3/3	267.33
1	Souvik Adhikary	3/3	277.80
3	Snehal Adhikary	3/3	263.01
4	Sumanth Andhavarapu	3/3	261.80
5	Kiran Hebasur	3/3	255.94
6	Samyucktha Ramesh	3/3	255.65
7	Perisetla Pavan Kalyan	3/3	255.39
8	SHASHANK JHA	3/3	253.86
9	Sri Harsha Nanduri	3/3	242.20
10	Rakesh Karade	3/3	238.79

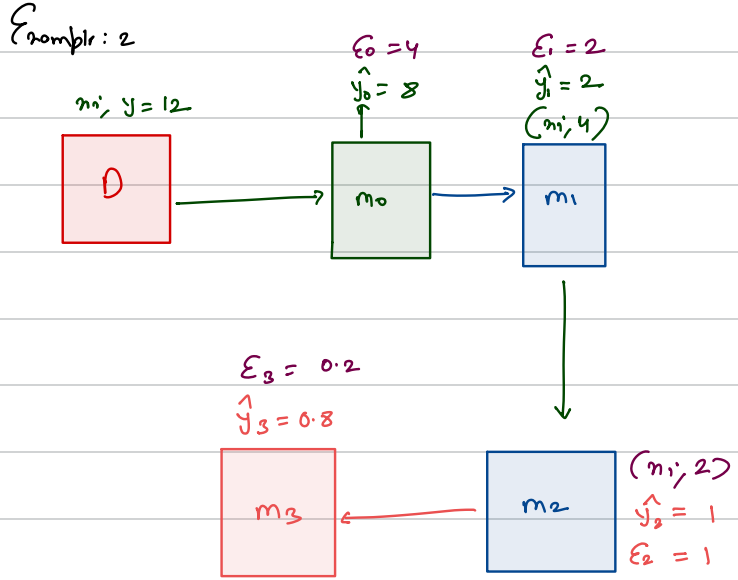
$M_0 \quad M_1 \quad M_2 \quad M_3 \quad - \quad - \quad - \quad .$

$h_1(x)$ = Prediction of $M_1(x)$ model only

$F_1(x) = \sum_{i=0}^{1} h_i(x) \Rightarrow h_0(x) + h_1(x)$

$h_3(x)$ = Prediction of $M_3(x)$

$F_3(x) = h_0(x) + h_1(x) + h_2(x) + h_3(x)$



$F_{pred}(x_i) = F_0(x_i) + \gamma \times \Pi_1(x_i) + \gamma \times \Pi_2(x_i) + \gamma \times \Pi_3(x_i)$

Need less

Effect of Learning Rate

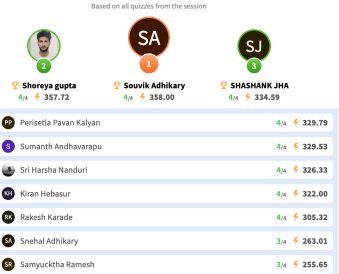
Why are we using DT with low depth here ?

0 users have participated

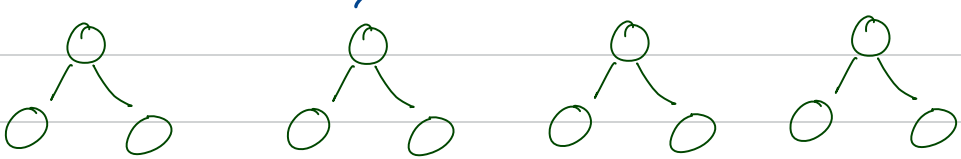
A	we are looking for high variance low bias mode	0%
B ✓	we are looking for high bias low variance model	0%
C	we are looking for overfit model	0%
D	no reason. just picked it randomly.	0%

End Quiz Now

Based on all quizzes from the session



Age, Gender



Π_1 Π_2 Π_3 Π_4

Complementing previous tree's weakness

y − h0(x) + h1(m) = F1(m)

What error shall we use here at stage 2 ?

0 users have participated

✓ A	Residual left after subtracting final model prediction (F_1(x)) from actual value	0%
B	Residual left after subtracting Stage 1 model (M_1) from actual value	0%
C	Residual left after subtracting Stage 0 model (M_0) from actual value	0%
D	None of the above	0%

End Quiz Now

Based on all quizzes from the session

2	Kiran Hebasur	5/5 386.39
1	Sri Harsha Nanduri	5/5 399.19
3	Souvik Adhikary	4/5 358.00
4	Shoreya gupta	4/5 357.72
5	SHASHANK JHA	4/5 334.59
6	Samyucktha Ramesh	4/5 331.45
7	Perisetla Pavan Kalyan	4/5 329.79
8	Sumanth Andhavarapu	4/5 329.53
9	Rakesh Karade	4/5 305.32
10	Snehal Adhikary	3/5 263.01

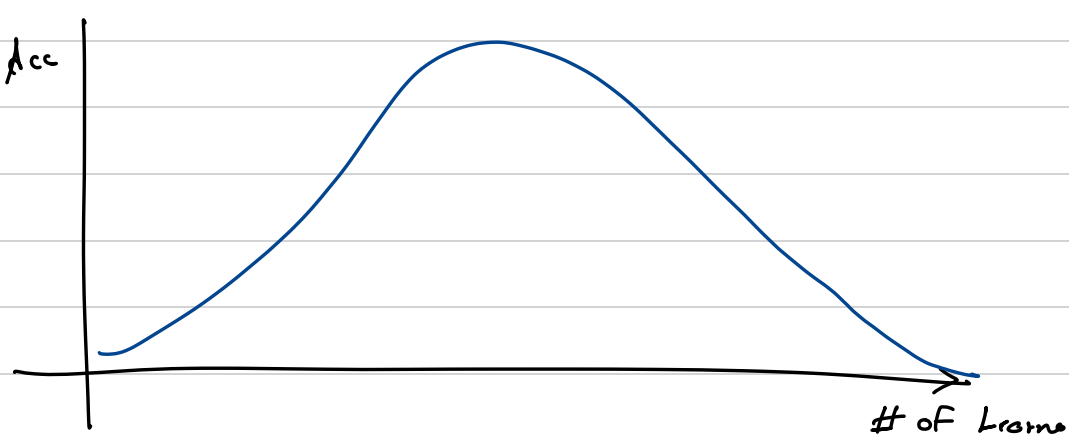
We keep doing this till **M stages**

$$F_m(x) = h_0(x) + \gamma_1 h_1(x) + \gamma_2 h_2(x) + \ldots \gamma_m h_m(x)$$

We can also write it as : $F_m(x) = h_0(x) + \sum_{i=1}^{m} \gamma_i h_i(x)$

Note:

M is a hyper parameter

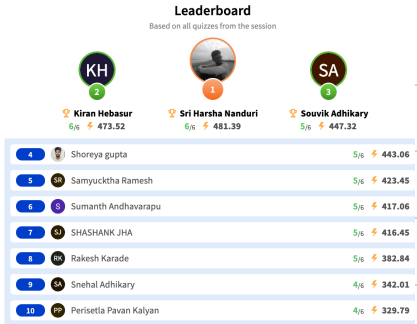


Which of the following statement is true?

0 users have participated

A	Boosting is a sequential model building technique	0%
B	We use weights to adjust the influence of model in final predicition	0%
C	New model is build on the error of previous models	0%
D ✓	All of the above	0%

End Quiz Now



Summarizing Process using an Example

Let's recap using the regression example i.e predict weight using height & gender.

Height	Gender	Weight(y)
1.6	M	82
1.5	F	55
1.4	F	61
1.4	M	65

STEP 0: Mean Model

$$h_0(x) = \bar{y}$$

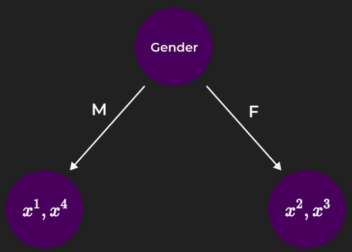
$$= \frac{82 + 55 + 61 + 65}{4}$$

$$h_0(x) = 65.75$$

Using the prediction of Stage 0 we calculate residual / error.

Height	Gender	Weight(y)	err_0
1.6	M	82	16.25
1.5	F	55	-10.75
1.4	F	61	-4.75
1.4	M	65	-0.75

Stage 1 : Using the residual (err_0^i) & features build a low depth DT



	Height	Gender	err_0	h1(x)
x1	1.6	M	16.25	7.75
x2	1.5	F	-10.75	-7.75
x3	1.4	F	-4.75	-7.75
x4	1.4	M	-0.75	7.75

$$\hat{y}_{left} = \frac{16.25 - 0.75}{2}$$

= 7.75

$$\hat{y}_{right} = \frac{-10.75 - 4.75}{2}$$

= - 7.75

Why is boosting called "Gradient Boosted Decision Trees"

Loss function = $(y_i - \hat{y})^2$ mse

$\frac{\partial L}{\partial \hat{y}_i}$ $2(y_i - \hat{y})$

gradient / Residual

$y_i - \hat{y} \approx$ Pseudo gradient

What do you think are the hyperparameter for boosting ?

0 users have participated

A	Number of trees	0%
B	Depth	0%
C	which loss to use	0%
✓ D	All of the above	0%

End Quiz Now

Based on all quizzes from the session

2	Sri Harsha Nanduri	7/7 565.53
1	Kiran Hebasur	7/7 567.39
3	Souvik Adhikary	6/7 543.25
4	Shoreya gupta	6/7 535.62
5	Samyuckta Ramesh	6/7 518.11
6	Sumanth Andhavarapu	6/7 513.16
7	Rakesh Karade	6/7 469.44
8	Snehal Adhikary	5/7 438.71
9	SHASHANK JHA	5/7 416.45
10	Tanvi Singh	5/7 355.17