Random Forest → Does not learn from mistake

AdaBoost → Focus on hard points

Gradient Boosting
↳ Learn from errors
→ measure how bad the error is
→ step by step improvement

* Our plan for this module:

i) Introduction to Gradient Boosting
ii) Limitation of AdaBoost & Learning from Errors
iii) Gradient Boosting Training, prediction & key parameters
iv) Implementation on a Dataset
v) Model Evaluation, Overfitting control & Use cases

Introduction to Gradient Boosting

Core Intuition:

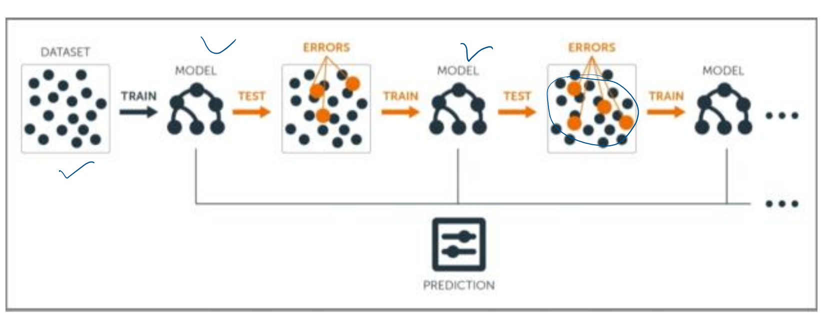
Each step adds a small correction that improves prediction quality.

Key variables:

x: input features
y: true target
$\hat{y}$: predicted target
$F(x)$: current model
$F_0(x)$: Initial model
m: boosting step index
$T_m(x)$: mth weak learner
η: learning rate

Model Form:

$$F_M(x) = F_0(x) + \eta T_1(x) + \eta T_2(x) + \cdots + \eta T_M(x)$$

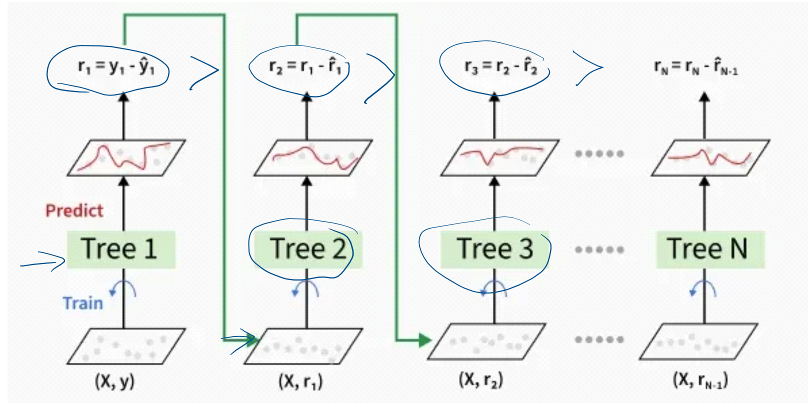


$T_m(x)$: mth weak learner

η: learning rate

$L(y,\hat{y})$: Loss function

r_{im}: residual



Step-by-Step Gradient Boosting Example

Loss Function:

Example:

Actual value = 10

Prediction = 7

Error = 3

Squared Loss:

$$\underbrace{L(y,F)}_{} = \frac{1}{2}(\underbrace{y-F}_{\text{Error}})^2$$

y: Actual value

F: model Prediction

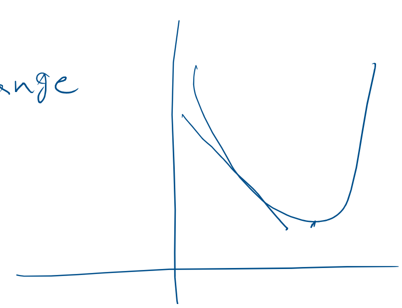
$1^2 = 1$

$5^2 = 25$

$10^2 = 100$

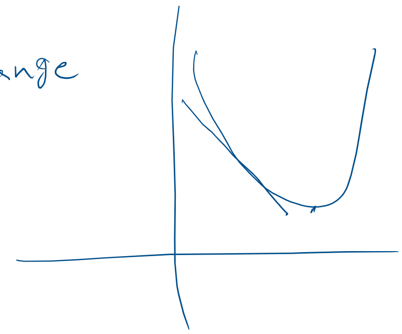
Derivative:

If Prediction changes then how the loss will change



Derivative:

If prediction changes then how the loss will change



Squared Loss Derivative

$$L(y,F) = \frac{1}{2}(y-F)^2$$

$$\Rightarrow \frac{\partial L}{\partial F} = -(y-F)$$

$$-\frac{\partial L}{\partial F} = -(-(y-F)) = y-F \rightarrow \text{Residual} = \text{Negative Gradient}$$

Residual:

$y = 10$

$F = 11$ or 8

$r_1 = 10-11 = -1 \rightarrow$ model predicted large value than actual

$r_2 = 10-8 = 2 \rightarrow$ model predicted less than actual

Tiny Dataset

i	x_i	y_i	$\hat{y}$	$r_{i1} = y - \hat{y}$	$F_1(x_i) = F_0(x_i) + \eta T_1(x_i)$	$r_{i2} = y_i - F_1(x_i)$
1	1	3	6	-3	5	-2
2	2	5	6	-1	5	0
3	3	7	6	1	7	0
4	4	9	6	3	7	2

Step 0:

Initialize $F_0(x)$

Step 0:

Initialize $F_0(x)$

$$F_0(x) = \bar{y} = \frac{3+5+7+9}{4} = 6$$

Initial prediction for all points:

$$\hat{y}^{(0)} = 6$$

Step 1:

Residual calculation

$$r_{i1} = y_i - F_0(x_i)$$

Step 2: Fit First weak Learner $T_1(x)$ on Residuals

x values: 1, 2, 3, 4

Midpoint splits: 1.5, 2.5, 3.5

$\Downarrow$

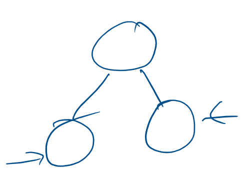
Left group: $x \le 2.5 \rightarrow$ residuals $= (-3, -1)$

Righ group: $x > 2.5 \rightarrow$ residuals $= (1, 3)$

Mean Residuals:

$$\text{Left mean} = \frac{-3 + -1}{2} = -2$$

$$\text{Righ mean} = \frac{1+3}{2} = 2$$



So, tree

$$T_1(x) = \begin{cases} -2, & x \le 2.5 \\ 2, & x > 2.5 \end{cases}$$

Step 3: Update model with new prediction

Learning rate,

$$\eta = 0.5$$

$$F_0(x) + \eta T_1(x)$$

$\eta = 0.5$

$F_1(x) = F_0(x) + \eta T_1(x)$

Case A: $x \leq 2.5$

$F_0(x) = 6$

$T_1(x) = -2$

$\eta = 0.5$

$F_1(x) = 6 + 0.5 \times (-2)$

$= 6 - 1 = 5$

if $x \leq 2.5$, then the new prediction is $= 5$

Case B: $x > 2.5$

$F_0(x) = 6$

$T_1(x) = +2$

$\eta = 0.5$

$F_1(x) = 6 + 0.5 \times (+2)$

$= 6 + 1 = 7$

if $x > 2.5$, then the new prediction is $= 7$

Previous residuals : [−3, −1, 1, 3]

current " : [−2, 0, 0, 2]

Key Learning:

1. To calculate residual
2. Fit the residual to the model (tree)
3. Controll update using learning rate
4. we check the residual again