Boosting

13 December 2022 20:38

Boosting

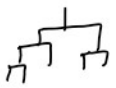
(Parallel)

Bagging :- Base models + aggregation

RF :- [Low bias. Base models (Deep trees (overfit) high variance)] + (Row sampling + column sampling) (diversify) + [Agg. (low variance)]

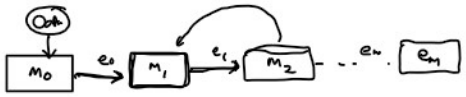
Boosting – (high bias) Base models + addition (sequence)

Shallow trees



underfit ✓

overfit

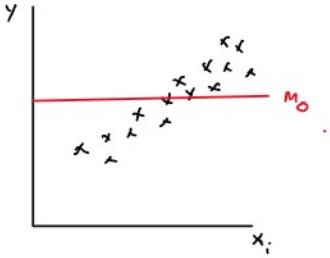


n obs (x_i, y_i)

x_1	x_2	y	e_0	$\hat{y}$
-	-	9	-7	16
		15	-1	16
		20	4	16
		20	4	16 $= M_1$

M_0 Step 0 (Base model)

M_0 – average value of y ($\bar{y} = 16$)



$\hat{y}_i : M_1 = h_1(x_1, x_2)$

Step 1

M_1 ← x_1 x_2

e_0	$\hat{y}_1$	e_1
-7	-3	-4
-1	0	-1
4	2	2
4	3	3

y	M_0 $\bar{y}$	e_0	$\hat{y}_1$	e_1
20	16	4	1	3

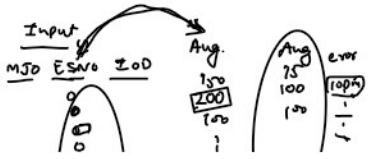
$M_1 = 16 + 1$

$M_1^1 = M_0 + M_1 = 16 + 1 = 17$

(50) → (25)

100 + 25 '125'

"EL NINO" "LA NINA"



$M = M_0 + M_1 + M_2 + M_3 + \cdots + M_m$

(x, y) (x, e_0) (x, e_1)

$$M = \underset{(x,y)}{M_0} + \underset{(x,e_0)}{M_1} + \underset{(x,e_1)}{M_2} + M_3 + \dots M_m$$

$\bar{y}$ $\quad \hat{y}_1 \quad \hat{y}_2$

$e_0 = y_i - \bar{y} \quad e_1 = e_0 - \hat{y}_1 \quad e_2 = e_1 - \hat{y}_2 \quad \dots \dots \quad e_m$?

$M_0 = \bar{y}$

$M_1 = \bar{y} + \hat{y}_1$

$M_3 = \bar{y} + \hat{y}_1 + \hat{y}_2$

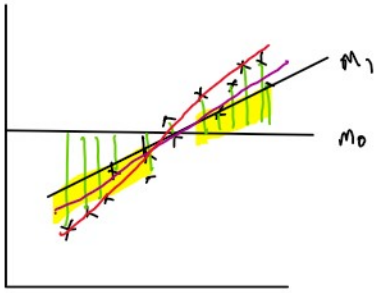
$$\underline{M} = \alpha_0 M_0 + \alpha_1 M_1 + \alpha_2 M_2 + \dots \alpha_m M_m$$

$$F(x) = \alpha_0 h_0(x) + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \dots \alpha_m h_m(x)$$

Gradient Boosting Trees

→ Regression ✓

→ classification



Gradient Boosting

Y	M_0	M_1	M (α_0, α_1)
9	16	-3	$\alpha_0 16 - \alpha_1 3$
15	16	0	$\alpha_0 16 + \alpha_1 0$
20	16	2	$\alpha_0 16 + \alpha_1 2$
20	16	1	$\alpha_0 16 + \alpha_1 1$

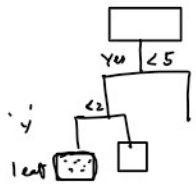
(MSE)

$$\frac{1}{n}\left[(9 - (\alpha_0 16 - \alpha_1 3))^2 + 15 - \right]$$

(α_0, α_1)

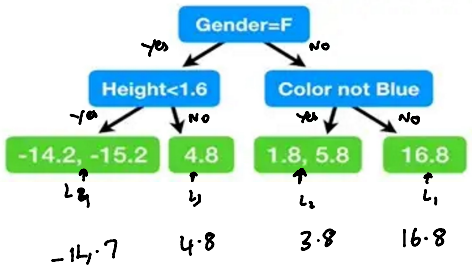
Step3 : $\alpha_0 \times h_0(x) + \alpha_1 \times h_1(x) + \alpha_2 \cdot h_2(x)$

Height (m)	Favorite Color	Gender	Weight (kg)	$\bar{y}$ = 71.2	e_0	$\hat{y}_1$
1.6	Blue	Male	88	71.2	16.8	16.8
1.6	Green	Female	76	71.2	4.8	4.8
1.5	Blue	Female	56	71.2	-15.2	-14.7
1.8	Red	Male	73	71.2	1.8	3.8
1.5	Green	Male	77	71.2	5.8	3.8
1.4	Blue	Female	57	71.2	-14.2	-14.7



Step-1 $\alpha_0 = 1, \alpha_1 = 1$

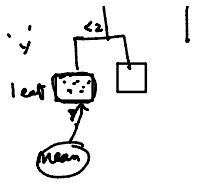
Step-1 Model



Step-1 $\alpha_0 = 1,\ \alpha_1 = 1$

$\alpha_0 \cdot m_0 + \alpha_1 m_1$

$71.2 + 16.8 = 88$



$\vartheta = 0.01$

Learning rate / Shrinkage

$M = \alpha_0 \cdot m_0 + \vartheta \cdot \alpha_1 m_1 + \vartheta\, \alpha_2 m_2 + \vartheta \cdot \alpha_3 m_3$

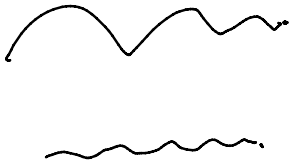
$\vartheta = 0.1$

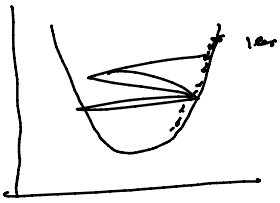
$\vartheta = 1$ ← (no. of models)

$M = \alpha_0 m_0 + 0.1 \cdot (\alpha_1 \cdot m_1)$

$= 71.2 + 0.1\,(16.8)$

$= 71.2 + 1.68 = 73$





$$\text{Model Error} = \text{Bias}^2 + \text{Variance} + \text{unexplainable error}$$

Bias ↓ low

Variance ↓ low

Balance b/w Bias and variance

Overfitting (Boosting) →

→ don't train too many Base models (n-estimators)

→ keep the learning rate or Shrinkage (low)

→ Base models → shallow (high Bias)

Tree Control parameters

→ max depth

→ min no. of Sample leaf node

→ max no. of leaf nodes

no. of estimator vs Learning rate

← (Small) (very minute steps)

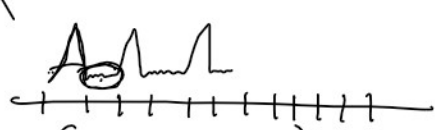
more no. of estimator ← large

len no. of estimators.

① More no. of estimator, and large learning rate?

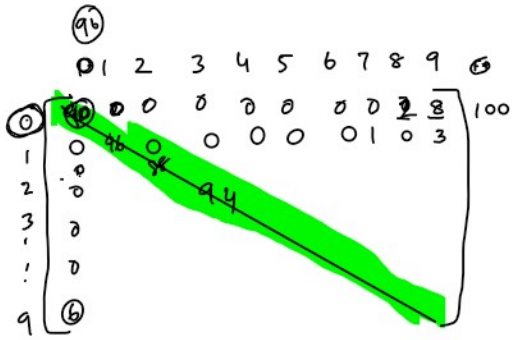
what happen? (Rectification)

Noise handling :- (Running) → (Brain Signals) ←



Denoising → (Moving average)

100/1000 Smoothens, Remove small Spikes (diffractions)



$\frac{90+96+\dots}{1000} = 0.95$

pen "0"

	0	for
0	90	10
for	6	2