GBM Regressor

Person	Height (cm)	Favorite Color	Gender	Weight (kg)
1	180	Blue	Male	88 − 71.2
2	165 ✓	Red	Female	76 − 71.2
3	175	Green	Male	80 − 71.2
4	160 ✓	Blue	Female	70 − 71.2
5	170	Red	Male	65 − 71.2
6	155 ✓	Green	Female	48 − 71.2

Pseudo residual

16.8
4.8
8.8
−1.2
−6.2
−23.2

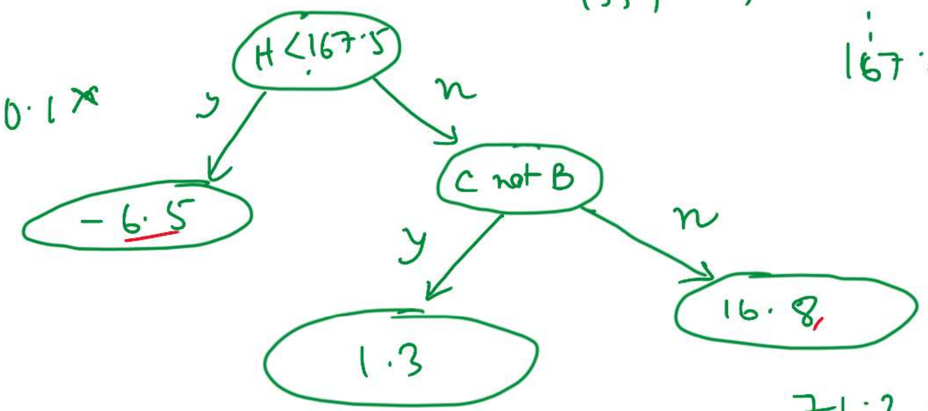
155, 160, 165, 170, 175, 180

167.5

Initial Prediction (leaf) $= \frac{1}{6} \times (88+76+80+70+65+48)$

$= 71.2$

$\boxed{71.2} + 0.1 \times$



$\frac{4.8 - 1.2 - 23.2}{3}$

$= -6.5$

$\frac{8.8 - 6.2}{2}$

$71.2 + 0.1 \times 16.8 = 71.2 + 1.68 = \boxed{72.88}$ ✓

$71.2 + 0.1 \times (-6.5)$

$= 71.2 - 0.65$

$= 70.55$ ✓

(MSE)

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

$$\underbrace{-\frac{\partial L}{\partial F}}_{\text{Gradient}} = (\underbrace{y}_{\text{observed}} - \underbrace{F}_{\text{predicted}}) = (\text{Pseudo - Residual})$$

(BCE)

$$L = -\left[y \log p + (1-y)\log(1-p) \right] \qquad p = \frac{1}{1+e^{-F}}$$

$$\frac{\partial L}{\partial p} = -\left[\frac{y}{p} - \frac{(1-y)}{(1-p)}\right] = -\left[\frac{y(1-p) - p(1-y)}{p(1-p)}\right]$$

$$= -\left[\frac{y - \cancel{py} - p + \cancel{py}}{p(1-p)}\right] = -\left[\frac{y-p}{p(1-p)}\right]$$

$$\Rightarrow -\frac{\partial L}{\partial p} = \frac{y-p}{p(1-p)} = \frac{\text{Residual}}{p(1-p)}$$

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

$$L = MSE + L2 = \frac{1}{2}(y-F)^2 + \frac{\lambda}{2}F^2$$

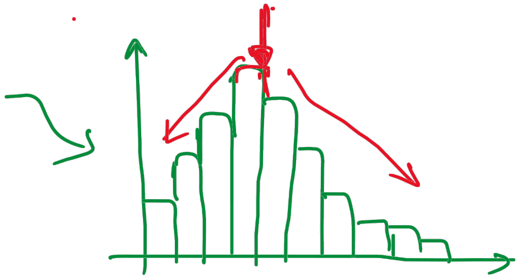
$$\therefore \frac{\partial L}{\partial F} = -(y-F) + \lambda F = -y + (1+\lambda)F = 0$$

$$\therefore F^* = \frac{y}{\underline{1}+\lambda} = \frac{y}{h+\lambda}$$

$$h \doteq \frac{\partial^2 L}{\partial F^2} = \frac{\partial}{\partial F}\left(-(y-F)\right)$$

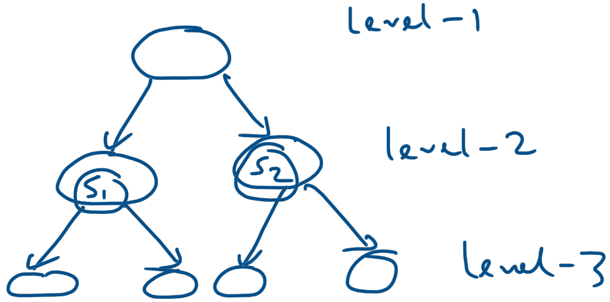
$$= \frac{\partial}{\partial F}(-y+F) = 1$$

Sorted values



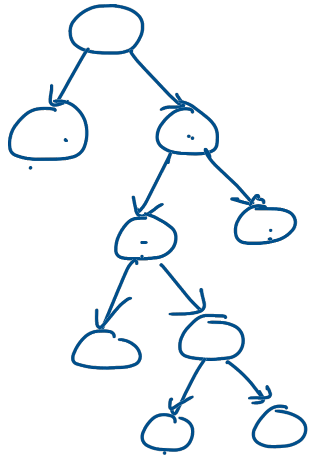
histogram.

Level wise growth :-



Leaf-wise growth

Best First Search



$\boxed{O(nd)}$

$O(n'd)$

$n' << n$ → sampled values

$\boxed{\sum \text{residuals}}$ →

	|residual|	Sort
1	10.	10
2	2	8
3	8	7
⋮	⋮	6
⋮	⋮	⋮
10		1

LR $(0.3) = a$

SR → 3 values $\left(\frac{3}{7}\right) = b$

GOSS

→ { Large residuals + Scaled samples of small residuals }

$$\frac{1-a}{b} = \frac{1-0.3}{\frac{3}{7}} = \frac{0.7}{3} \times 7 = \frac{4.9}{3} \approx \boxed{1.6}$$

Colour (x)	target (y)	target Encoding
Blue	0	0.33
Red	1	0.67
Red	1	0.67
Blue	0	0.33
Red	0	0.67
Blue	1	0.33

$x =$ 'Blue' , $y = \{0, 0, 1\}$

$x =$ 'Red' , $y = \{1, 1, 0\}$

$$\underset{(x=c)}{\text{target encoding}} = \frac{\sum_{x=c} y}{\sum \mathbb{1}(x=c)}$$

$$(x = \text{'blue'}) = \frac{0+0+1}{3} = 0.33$$

$$(x = \text{'Red'}) = \frac{1+1+0}{3} = 0.67$$

target - leakage problem $\rightarrow$ overfitting

Ordered Target Statistics (CatBoost) : Avoids target leakage

→ for each sample, compute mean target value using only samples that come before it in a random permutation of dataset.

$$TS_k(x) = \frac{\sum_{j<k,\, x_j = x} y_j + a \cdot P}{\sum_{j<k,\, x_j = x} 1 + a}$$

a → smoothing parameter

P → prior (global average)

Ex:-

Row	Colour	target
1	Red	1
2	Blue	0
3	Red	0
4	Blue	1
5	Red	1

P = Global average (prior) $= \frac{\sum y_j}{\text{Count}(y_j)} = \frac{3}{5}$

$= 0.6$

Suppose $a = 1$

Permutation-1 : [3, 1, 5, 2, 4]

Permuted dataset

	Row	Colour	y	target Enc[1]
$j=0 \rightarrow$	$\boxed{3}$	$\boxed{R}$	0	$\frac{0+1\times 0.6}{0+1} = 0.6$
$\rightarrow$	1	R	1	$\frac{0+1\times 0.6}{1+1} = 0.3$
$j=2 \rightarrow$	5	$\rightarrow$ R	1	$\frac{0+1+1\times 0.6}{2+1} = \frac{1.6}{3} = 0.53$
	2	B	$0 \rightarrow 0.6$	
	4	B	$1 \rightarrow 0.3$	

Initial data

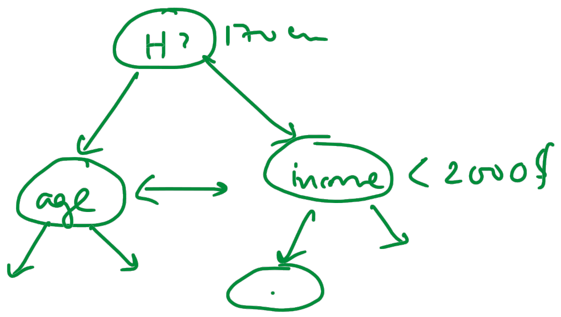
Row	Colour	y	ts_1	ts_2	ts_3	⟨TS⟩
1	R	1				
2	B	0				
3	R	0				
4	B	1				
5	R	1				

$$TS(R)_{[j=0]} = \frac{0+1\times 0.6}{0+1} = 0.6$$

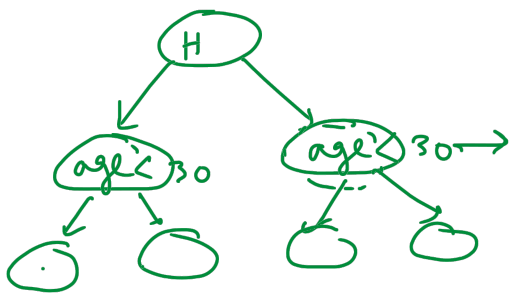
$$TS_k(x) = \frac{\sum_{j<k,\, x_j = x} y_j + a \cdot P}{\sum_{j<k,\, x_j = x} 1 + a}$$

$$TS(R)_{[j=2]} = \frac{(0+1) + 1\times 0.6}{2+1} = \frac{1+0.6}{3} = \frac{1.6}{3} = 0.53$$

$$TS(R)_{[j=1]} = \frac{0+1\times 0.6}{1+1} = \frac{0.6}{2} = 0.3$$



normal tree
(normal boosting)



Github

Repo → Logicmojo-AI-Datascience-Assignment

→ Assignment-1

→ " -2

→ " -3

⋮