Gradient Boosting

① Additive Modelling

$X \{ y \longrightarrow \quad y = f(x)$

$y = f(x_1, x_2, x_3)$

Sum of Smaller function called additive modelling

$f(x) = X + \sin(x)$

Step:- Gradient Boosting

Input training set $\{(x_i, y_i)\}_{i=1}^{n}$, a Differentiable loss function $L(y, f(x))$, Number of iteration M.

Step 1. Initialize $f_0(x) = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i - \gamma)$

Step 2. For $m = 1$ to M

a. For $i = 1, 2 \ldots\ldots N$ compute

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\delta f(x_i)}\right]_{f = f_{m-1}}$$

b. Fit a regression tree to the targets r_{im} giving terminal regions.

$R_{jm}, \; j = 1, 2 \ldots\ldots, J_m$

c. For $j = 1, 2, \ldots, J_m$ compute

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$$

d. update

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$$

Step. 3. Output $\hat{f}(x) = f_M(x)$

Solve. 1- Step. I

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, \gamma)$$

Loss Function - least Square

$L(y, f(x))$

$L(y_i, \hat{y})$

$$L = \sum_{n=1}^{n} (y_i - \hat{y}_i)^2$$

$$L = \frac{1}{2} \sum_{n=1}^{n} (y_i - \hat{y}_i)^2$$

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} \cancel{\tfrac{1}{2}} (y_i - \hat{y}_i)^2$$

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \frac{1}{2} \sum_{i=1}^{n} (y_i - \gamma)^2$$

$$\frac{d f_0(x)}{d\gamma} = \frac{d}{d\gamma} \frac{1}{2} \sum (y_i - \gamma)^2$$

$$= \frac{-2}{2}\sum_{i=1}^{n} (y_i - \gamma) \quad \text{— Divide by } -1 \text{ both side}$$

$$= \sum_{i=1}^{n} (y_i - \gamma) = 0$$

$$= \sum_{i=1}^{n} (\gamma - y_i) = 0$$

$$= \sum (\gamma - y_1) + (\gamma - y_2) + (\gamma - y_3)$$

$$0 = -3\gamma + \sum y_1 + y_2 \; y_3$$

$$\boxed{\gamma = \frac{\sum y_i}{n}}$$

$$f_0(x) = \gamma$$

Step 2: a. for $i = 1, 2 \ldots N$ compute

$$r_{im} = -\left[\frac{\delta L(y_i \; f(x_i))}{\delta(f(x_i))}\right]_{f = f_{m-1}}$$

where i = row.

m = Decision tree model

residual.

Psudo residual for $m = 1$ $\quad r_{11}$

$$r_{11} = -\left[\frac{\delta L((y_i), f(x_i))}{\delta \; f(x_i)}\right]_{f = f_0}$$

Suppose we have 3 row in our Dataset we calculate.

Decision tree

r_{11} r_{21} r_{31} ---

$\hat{y}_i = f(x)$

$$r_{i1} = -\left[\frac{\delta L(y_i, \hat{y})}{\delta(\hat{y}_i)}\right]_{f=f_0}$$

$$L = \frac{1}{2} \sum (y_i - \hat{y}_i)^2$$

$$r_{i1} = -\left[\frac{\delta}{\delta \hat{y}} \frac{1}{2} (y_i - \hat{y}_i)\right]_{f=f_0}$$

Differentiated

$$r_{i1} = -\frac{1}{\cancel{2}} \times \cancel{2} (y_i - \hat{y}_i) = \left[(y_i - f(x_i)\right]_{f=f_0}$$

$$r_{i1} = (y_i - f_0(x_i))$$

$r_{11} = (y_1 - f_0(x_1))$

$r_{21} = (y_2 - f_0(x_2))$

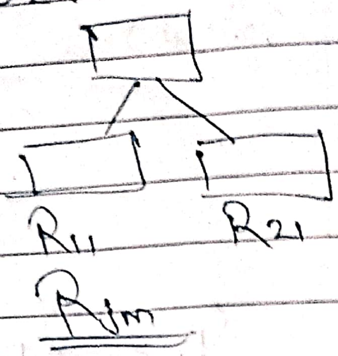
$r_{31} = (y_3 - f_0(X_3))$

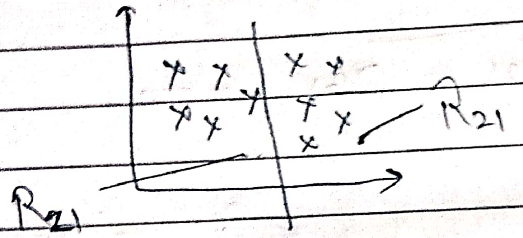
Step 2 b

Train Decision tree of Data where X columns are are Given Data & y is r_{i1}

Min Depth is 8 & Max is 32.

For Now. Consider Depth is 3 (1)





Step 2:- C

Calculate the output value of Gradient Boost

γ gamma value calculate

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$$

γ_{11} γ_{21} — R_{11}

In ower Decision tree we have two leaf node.

$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{11}} L(y_i, F_{m-1}(x_i) + \gamma)$$

$$\gamma_{11} = \underset{\gamma}{\operatorname{argmin}} \frac{1}{2} (y_i - F_{1-1}(x_i) + \gamma)^2$$

$$\frac{dL}{d\gamma} = \frac{1}{\cancel{2}} \times \cancel{2} (y_i - f_0(x) + \gamma) \frac{d}{d\gamma}(y_i - f_0(x) - \gamma)$$

$$= -(y_i - f_0(x) - \gamma)$$

$$\gamma_{11} = y_i - f_0(x) - \gamma = 0$$

$$\boxed{\gamma = y_i - f_0(x)}$$

Date / /

Now. leaf node have more than one Sample

$$\gamma_{21} = \underset{}{\arg\min} \sum_{x_i \in R_{21}} L(y_i, f_0(x) + \gamma)$$

$$= \underset{\gamma}{\arg\min} \frac{1}{2} \sum_{i=1}^{2} (y_i - f_0(x_i) + \gamma)^2$$

$$D/D\gamma = -\sum_{i=1}^{2} (y_i - f_0(x_i) - \gamma) = 0$$

$$\boxed{\gamma = \sum_{i=1}^{2} \frac{y_i - f_0(x_i)}{n}}$$

Step 2 :- d

$f_1(x) = f_0(x) +$ Decision tree output

$f_2(x) = f_1(x) +$ DT2 output.

$f_3(x) = f_2(x) +$ DT3 ot