* Gradient Boosting Regression

→ All boosting algorithm uses Additive Modelling concept. (stage wise additive modelling).

* Additive Modelling *

→ Machine learning is nothing but relationship between input and output, that relationship can be defined as mathematical function.

$$y = f(x), \text{--etc}$$

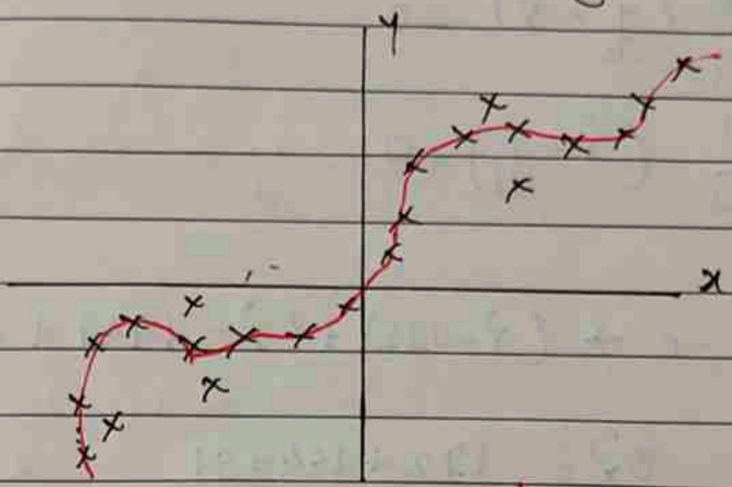
→ Sometimes relationship functions are very simple such as linear regression, polynomial regression etc.

→ But sometimes relationship functions are complex and composite and difficult to find out.

→ Breaking down such composite function into small function such that by addition of these small function we can get the composite function. Such method is called additive modelling.

→ Since we are adding the small functions in the steps it is called as stage wise modelling.

e.g



$y = x + \sin x \rightarrow y = f(x) \quad , \; y = f(\sin x)$

* Algorithm:

input: training set $\{(x_i, y_i)\}_{i=1}^{n}$ a differentiable loss function $L(y, F(x))$, number of iterations M.

$n = 3$

Loss function $= L(Y, F(x)) \longrightarrow L(Y, \hat{Y})$

Least Square $= \frac{1}{2} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$

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

$$f_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

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

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

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

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

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

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

for dataset

$$\sum_{i=1}^{3} (\gamma - y_i) = 0 \rightarrow (\gamma - 192) + (\gamma - 144) + (\gamma - 91) = 0$$

$$3\gamma = 192 + 144 + 91$$

$$\boxed{\gamma = \frac{192 + 144 + 91}{3}}$$

mean

$f_m(x) = f_0(x) =$ mean of output

Hence, proved 1st model is mean of y_i

$$f(x) = \underbrace{f_0(x)}_{\text{mean leaf}} + \underbrace{f_1(\gamma)\ f_2(\gamma) + \ldots + f_m(\gamma)}_{\text{same model}}$$

2\. Step 2: For $m = 1$ to M: → (Repeat steps M times)

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

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

$r \rightarrow$ residual

$i \rightarrow$ row number

$m \rightarrow$ Decision Tree Number

$$r_{i1} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_0}$$

3 rows: r_{11}, r_{12}, r_{31}

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

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

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

$\leftarrow \hat{y}_i = f(x_i)$

$$= \left[y_i - f(x_i)\right]_{f=f_0}$$

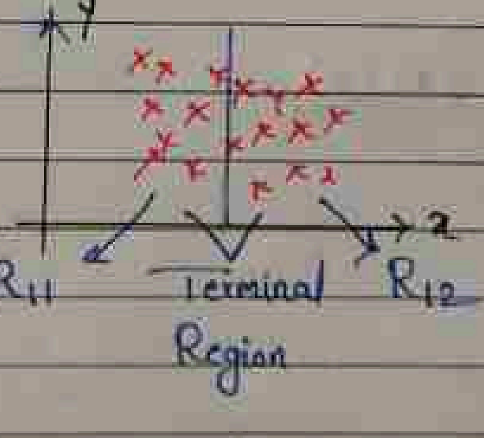
$$\boxed{r_{i1} = y_i - f_0(x_i)}$$

$r_{11} = y_1 - f_0(x_1) = 109 - 122$

$r_{21} = y_2 - f_0(x_1) = 101 - 122$

$r_{31} = y_3 - f_0(x_3) = 156 - 122$

b) Fit a regression tree to the targets r_{im} giving terminal regions R_{jm}, $j = 1, 2, \ldots J_m$.



c) For $j = 1, 2, \ldots, J_m$ Compute

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

$$\gamma_{11} = \arg\min_{\gamma} \frac{1}{2}(y_i - (f_0(x_i) + \gamma))^2$$

$$\frac{dL}{d\gamma} = \frac{1}{2} \times 2\left[y_i - (f_0(x) + \gamma)\right] \frac{d}{d\gamma}\left(y_i - f_0(x) - \gamma\right) = 0$$

$$-\left(y_i - f_0(x) - \gamma\right) = 0$$

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

$\gamma_{31} = 156 - 122 = 34$

γ_1

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

$$f_1(x) = f_0(x) + DT \rightarrow \text{output}$$

$$f_2(x) = f_1(x) + DT2 \qquad f_{(3)}(x) = f_2(x) + DT3$$

$$f_0(x) + DT \qquad F_1(x) + DT2$$

$$f_0(x) + DT1$$

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