Gradient Boosting Algorithm

Page No.
Date

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

	R&D Spend	Administration	Marketing Spend	Profit
0	165.0	137.0	472.0	192.0
1	101.0	92.0	250.0	144.0
2	29.0	127.0	201.0	91.0

training set

x_i — x_1, x_2, x_3

y_i — y_1, y_2, y_3

- Differentiable Loss function $= L(y, f(x))$, with $f(x) = \hat{y}$

A mathematical function tell How much mistake a algorithm is doing

+ we are taking Least Squared loss function

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

~~Step (1)~~

~~1~~ End goal

- we want to create a model b/w x and y that model is the relationship b/w x and y meaning we have to find a function of x on y

$y = f(x)$ → we are doing Boosting that why $f(x)$ will be composite of smaller function

$h(x) = h_0(x) + h_1(x) + f_2(x) \ldots\ldots f_n(x)$

$$f(x) = f_0(x) + [\underbrace{f_1(x) + f_2(x) \ldots f_n(x)}_{\text{Decision tree}}]$$

$\boxed{\text{Step 1}}$

$$\text{Initialize } f_0(x) = \underbrace{\arg\min_\gamma L(y_i, \gamma)}_{\text{formula}}$$

$$f_0(x) = \arg\min_\gamma L(y_i, \gamma) \quad \text{gamma}$$

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

Meaning :- ~~[illegible]~~
we want to find the value of γ for which the value of Loss function should be minimum

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

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

$$= \sum (y_i - \gamma) \frac{d}{d\gamma} (y_i - \gamma)$$

$$= -\sum (y_i - \gamma) = 0$$

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

Page No.
Date

putting our data set y_i values in $f_0(x)$

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

$$\sum_{i=1}^{3} (\gamma - y_i) = 0$$

$$= (\gamma - 192) + (\gamma - 144) + (\gamma - 91) = 0$$

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

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

meaning of our $f_0(x)$ is the mean of y_i

$\boxed{\text{Step } 2}$

for $m = 1$ to M:

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

residual

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

delta

i = row num

which model num

$\boxed{\text{For 1st iteration}}$ $\gamma_{i1} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f = f_0}$

So, in our dataset we only have 3 rows so we have to calculate r_{11}, r_{21}, r_{31}

Residual of row 1 for 1st Decision tree

Residual of row 2 for 1st Decision tree

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

As we know $f(x_i) = \hat{y}_i$

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

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

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

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

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

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

$$r_{11} = y_1 - f_0(x_1) = 197 - 142$$

$$r_{21} = y_2 - f_0(x_2) = 144 - 142$$

$$r_{31} = y_3 - f_0(x_3) = 91 - 142$$

Regidoal for 1st Decision Tree.

Step ②
(b) fit a regression tree to target r_{im} giving terminal regions R_{jm}, $j = 1, 2 \ldots J_m$

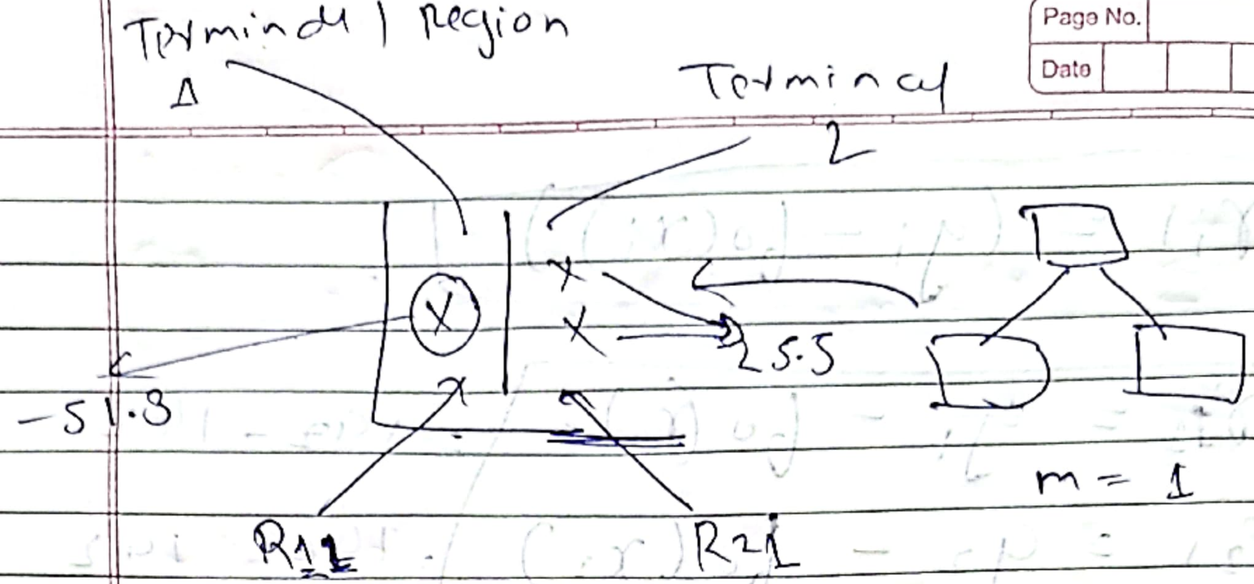
→ from sklearn.tree import DecisionTree Regressor
dt1 = DTR()

dt1.fit (x1, Regidoal for 1st DT)

Step ②
(c) for $j = 1, 2, \ldots J_m$ compute

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

calculating Output value for every terminal.

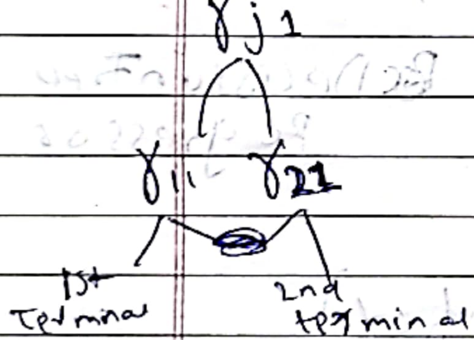


1st DT $\qquad\qquad$ 1st DT

1st Terminal $\qquad\qquad$ 2nd Terminal

So for calculating output value of every Terminal in Gradient Boosting. formula is

$$\rightarrow \gamma_{jm} = \underset{\gamma}{\arg\min} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{11} = \underset{\gamma}{\arg\min} \frac{1}{2} (y_i - (f_0(x_1) + \gamma))^2$$

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

$$\frac{d}{d\gamma} = (y_i - f_0(x) - \gamma) = 0$$

Page No.
Date

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

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

$= y_i - F_0(x) - \gamma = 0$

$\gamma_{11} = 91 - 142 - \gamma = 0$

$$\boxed{\gamma = 91 - 142 = -51}$$

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

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

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

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

$$= y_1 - f_0(x_1) - y + y_2 - f_0(x_2) - \gamma = 0$$

$$= 192 - 142 - \gamma + 144 - 142 - \gamma = 0$$

$$= 336 - 284$$

$$= 52 - 2\gamma = 0$$

$$= \gamma = \frac{52}{2} = 26$$

Page No.
Date

Step ②

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

D.T Output

So for

$f_1(x) = f_0(x) + dT$

$f_2(x) = f_1(x) + dT2$ output

$f_3(x) = f_2(x) + dt3$ output

$f_4(x) = f_1(x) + f_2(x) \ldots f_n(x)$

③ Output $f(x) = f_m(x)$