Gradient Boosting Maths:

Exp	Degree	Salary	$\hat{y}$
2	BE	50K	60
3	PHD	60K	60
4	MASTER	70K	60

residual

regression. $\boxed{Loss = \sum_{i=1}^{n} \frac{1}{2}(y-\hat{y})^2}$

we need to find $\hat{y}$ value such that, this summation should go down, loss function should go down. (very minimum)

consider above example,

$= \frac{1}{2}(50-\hat{y})^2 + \frac{1}{2}(60-\hat{y})^2 + \frac{1}{2}(70-\hat{y})^2$

first order derivative $\boxed{x^n \text{ derivate of } x = n x^{n-1}}$

$= \frac{2}{2}(50-\hat{y})(-1) + \frac{2}{2}(60-\hat{y})(-1) + \frac{2}{2}(70-\hat{y})(-1)$

$= -50+\hat{y} - 60+\hat{y} - 70+\hat{y}$ $\boxed{= \hat{y}}$ for Base.

$= -180 + 3\hat{y}$

$= 3\hat{y} = 180$ $\boxed{\therefore y = 60}$ = $\boxed{\text{constant value}}$

$Loss = \frac{1}{2}(y-\hat{y})^2 = \frac{2}{2}(y-\hat{y})(-1)$

$\partial L / \partial \hat{y} = -y + \hat{y} = -(y-\hat{y})$. $-\frac{\partial L}{\partial \hat{y}} = (y-\hat{y})$

Instead of gamma, we have $F_{m-1}(x_i) + \gamma$

$F_{m-1}(x_i)$ = previous model output.

To find γ ($\bar{y}$) to minimize the loss.

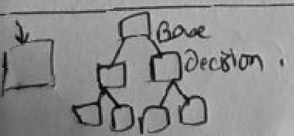
what is this loss,

$\sum_{i=1}^{n} \frac{1}{2}\left(y_i - (60 + \gamma \alpha(\bar{y})\right)^2$

find derivate.

(0.1) = learning rate

$60 + (0.1 \times 10) = 60 - 1.0$
$= 59.0$.



Pseudo algorithm

① Input $\boxed{\{x_i, y_i\}}$
Independent, dependent feature

② Loss function : $\boxed{L(y, F(x))}$

Different for regression, classif^n.

regression : Root mean square error, mean square error.

all the loss function should be differentiable. (derivate of loss funt^n)

classif^n : log loss, Hinge Loss

③ How many decision we require,

① $\boxed{\text{Initialize Model with constant value.}}$

~~$F_0(x) = Gamma(\sum ... (y$~~

$$\boxed{F_0(x) = \underset{\downarrow}{\arg\min} \left(\sum_{i=1}^{n} L(y, \gamma)\right) \quad =? \ \gamma \ (\text{Gamma})}$$

or $\hat{y}$.

Gamma = predicted value

② $\boxed{\text{Iterate } M = 1 \text{ to } M}$ M = no. of trees.

from 1 to M, I will compute pseudo residual, (pseudo error).

$$r_{im} = -\left[\frac{\partial L(y, F(x_i))}{\partial F(x_i)}\right]$$

error based on salary $-\hat{y}$

ⓐ Create DT, error = dependent, feature = Independent

Fit a base learner $h_m(x)$

$\boxed{i/p \{x_i, r_{im}\}}$ r_{im} = error value.

ⓑ $\boxed{\gamma_m = \underset{\gamma}{\arg\min} \sum_{i=1}^{n} L(y_i, F_m(x_i) + \gamma)}$

compare with initial model

ⓒ updating Model = $\boxed{F_m(x) = F_{m-1}(x) + L(h(x))}$