~~Boosting~~ :- Gradient Boosting Algo :-

↳ ~~ADA~~ → "Stepwise Additive modeling"

→ $h_i(x)$

↳ weak Model

→ $$F(x) = \sum_{i=1}^{M} \gamma_i \, h_i(x)$$

$F(x)$ ↓ Strong Model; γ_i ↳ weight; $h_i(x)$ ↳ weak model

→ $$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x).$$

$F_m(x)$ ↓ Final Model; $F_{m-1}(x)$ ↓ Previous Model; γ_m ↓ weighted; $h_m(x)$ ↓ Current Model

→ $$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_m(x)).$$

In Case of Regression

$$L(\omega) = \frac{1}{2N} \sum_{n=1}^{N} (y_n - (X\omega)_n)^2$$

$$\cancel{\frac{\partial L}{\partial \omega_j} = \frac{1}{N} \sum_{n=1}^{N} x_{nj} (\ )}$$

$$\frac{\partial L}{\partial \omega_j} = -\frac{1}{N} \sum_{n=1}^{N} (y_n - (X\omega)_n) \, x_{nj}$$

→ $F_0(X) = h_0 = \text{mean}(y)$ (base Model)

$\epsilon_1 = y - F_0(X)$.

$F_1(X) = F_0(X) + h_1(\epsilon_1)$.

$\epsilon_2 = y - F_1(X)$

$F_2(X) = F_1(X) + h_2(\epsilon_2)$

Until it start overfitting or sum of residual become constant, Overfitting can be checked accuracy on validation data.

Predictions = y_i^p

Loss = $L(y_i, y_i^p)$

Loss = MSE = $\Sigma (y_i - y_i^p)^2$

① mu = mean (y) → start with base learner

② dY = y − mu

③ for k=1 to nboost

learner (k) = train_Regressor (x, dY)

alpha (k) = 1 → learner rate

dY = dY − alpha (k) * predict (learner (k), x)

④ For test data

[Ntest, D] = Size of test data.

predict = zeros [Ntest, 1);

⑤ for k=1 to nboost

predict = predict + alpha (k) * predict (learner(k), Xtest)