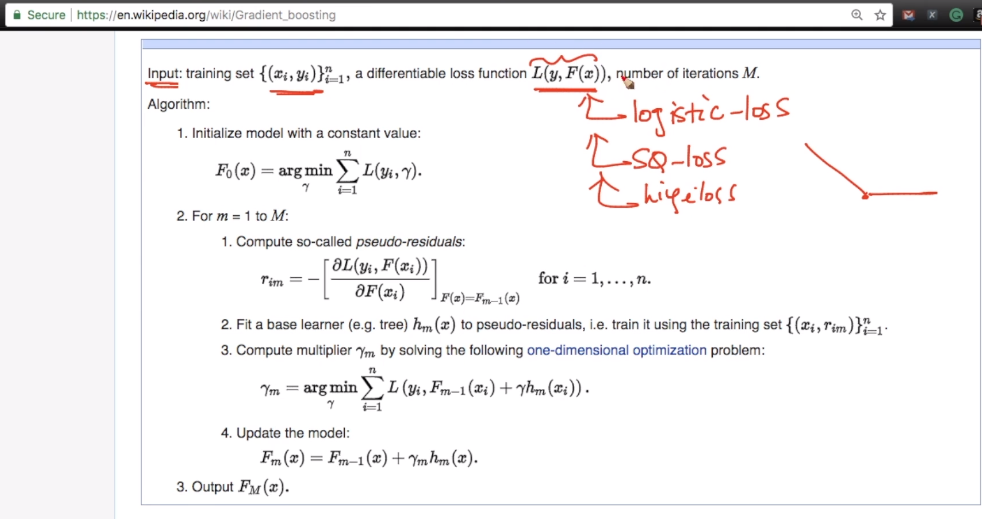
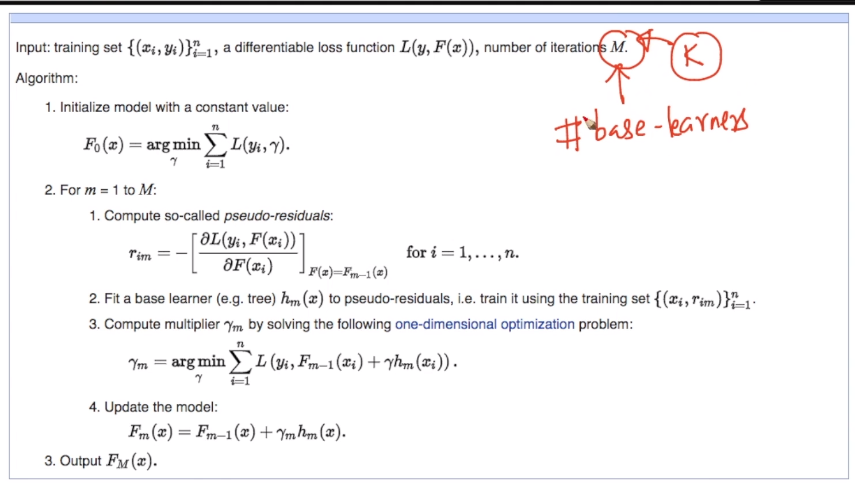
**Gradient Boosting**

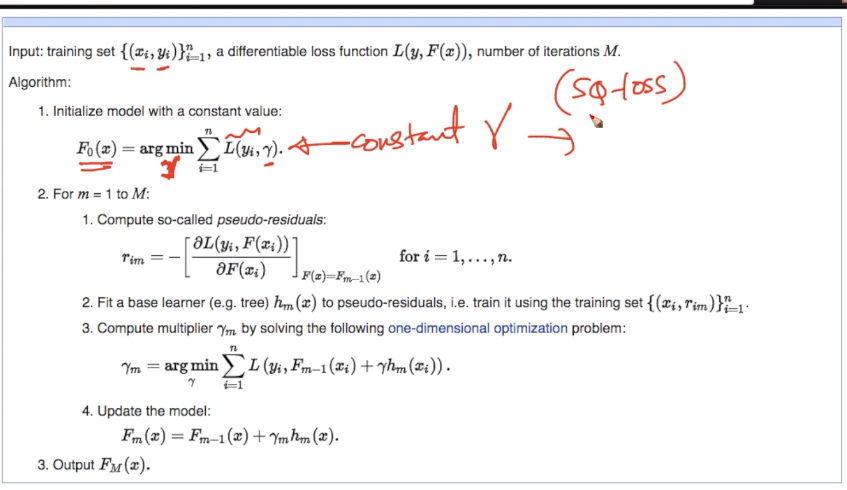
Below figure shows algorithm of gradient boosting

In this algo we provide training set, differentiable loss function L (it can be any loss function logistic loss, squared loss, hinge loss(although hinge loss is not differentiable at hinge itself but there is approximation which takes care of it)) and M which is no. of base learners

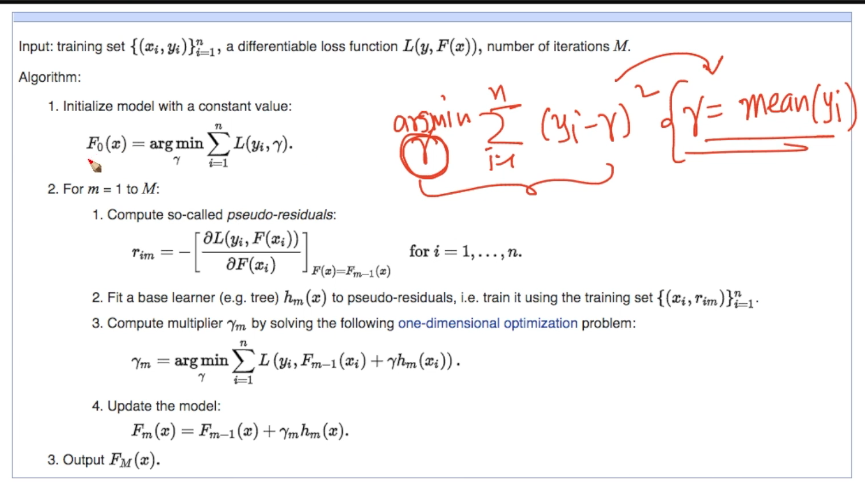




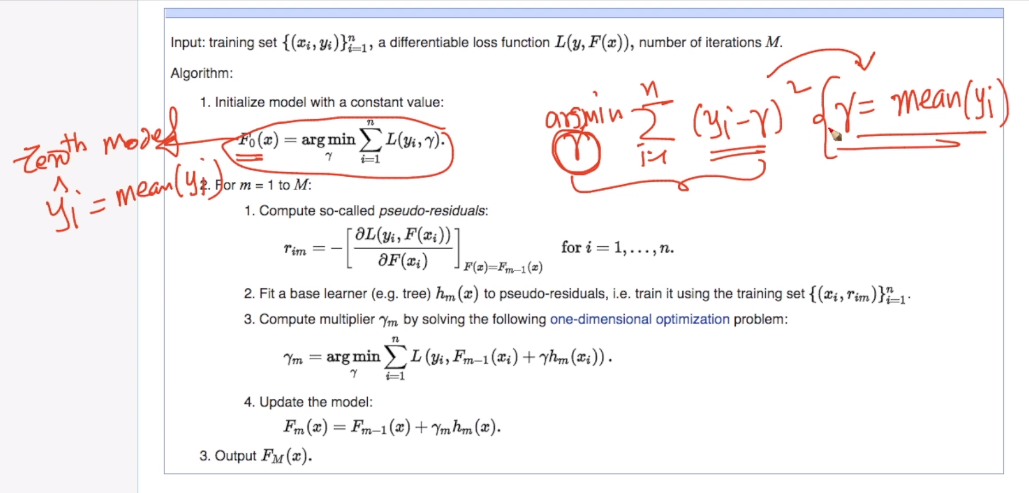
We have to initiate model M0(zeroth model) with constant gamma which should be minimum.



And in squared loss minimum gamma got at mean(yi)

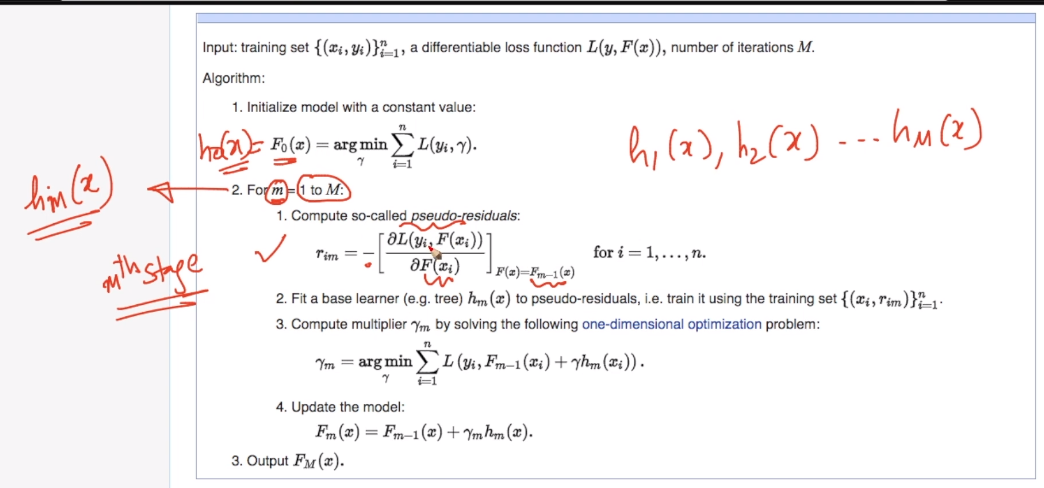


In zeroth model predicted yi  is mean(yi) for squared error loss

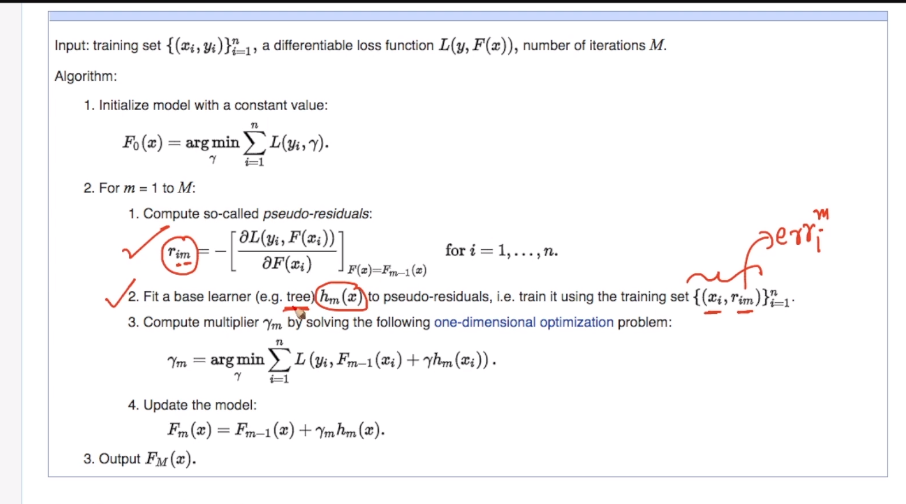


Now we iterate M times i.e no. of base learners in this

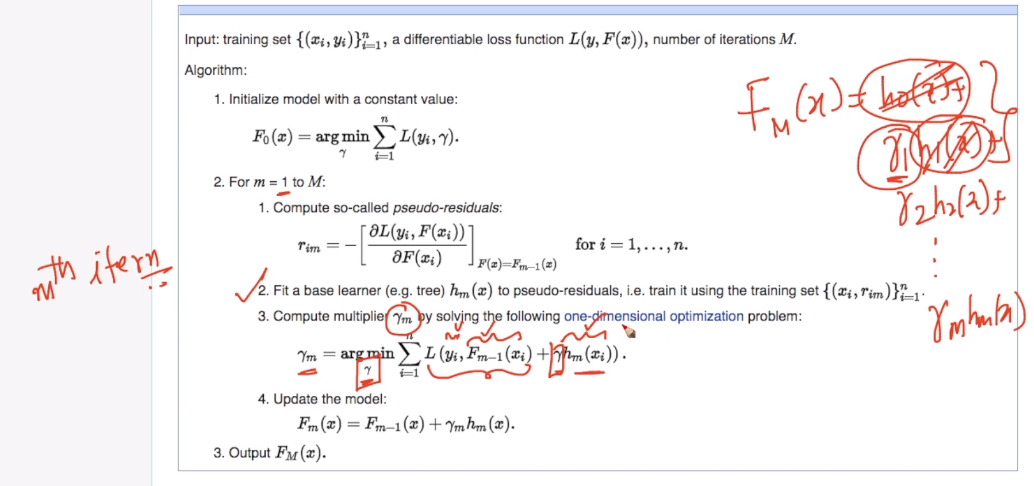
First we calculate pseudo residuals rim (pseudo residual error at mth stage for ith point is rim) which nothing but derivative of loss function



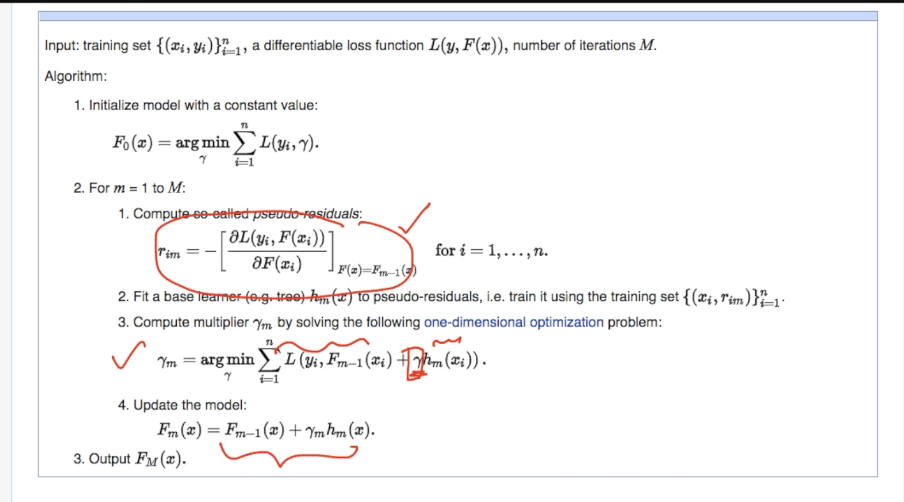
Now we fit base learner hm(x)



Now compute minimum gamma which can be done easily as we have all values in equation except gamma.



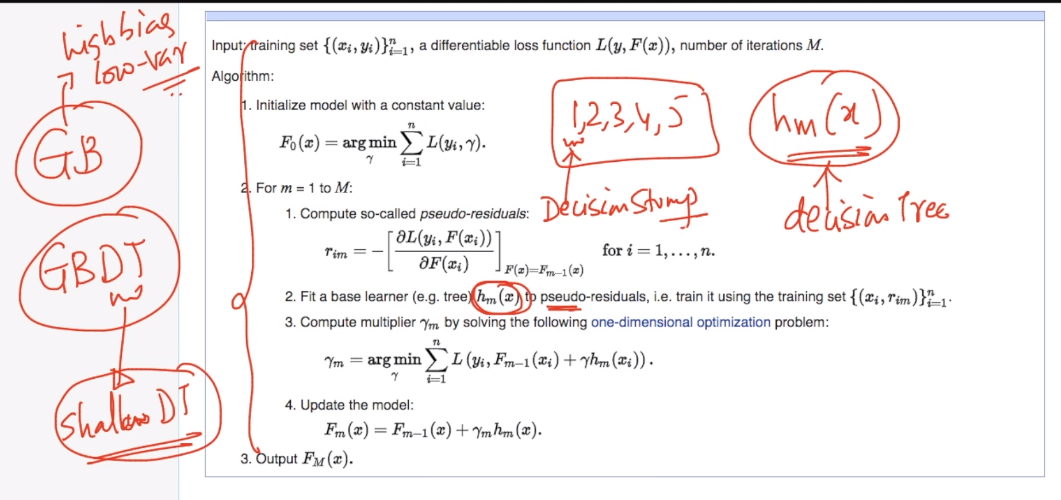
Now at final stage 4 we just sum up as shown below



Now the thing is which base learner we have to use .

GB is general idea for high bias and low variance model but GBDT is popular idea. And in decision tree we got high bias when we have shallow depth

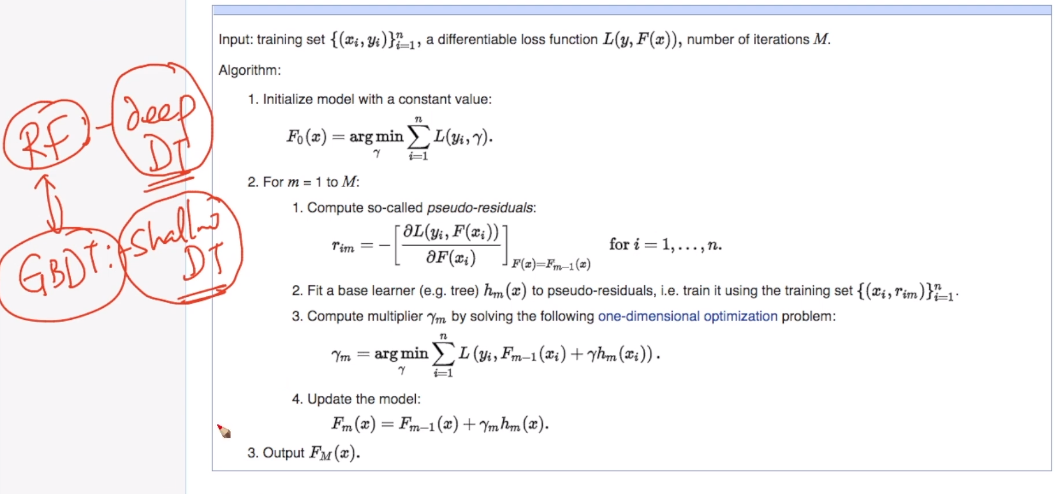
Decision tree with depth 1 is also know as decision stump and most of the times it work well.



So difference Between RF and GBDT is :

RF : in this we use low bias high variance model i.e decision tree with high depth

GBDT : in this we use high bias low variance model i.e decision tree with shallow or low depth



Comments:

