What is Boosting?

In Boosting,

- we build a bunch of simple models and
- Each model is trained using the residual of the previous model.
- use these models to build an additive weighted model

Base learners typically have low variance and high bias

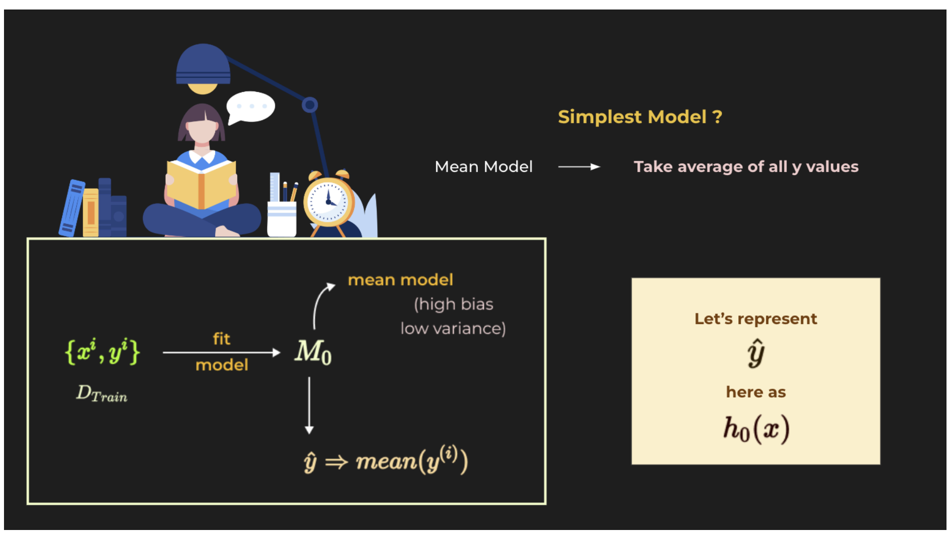
What sort of DT models have high bias?

- Shallow Trees or Decision Stump

The output of these models is combined in an Additive Manner

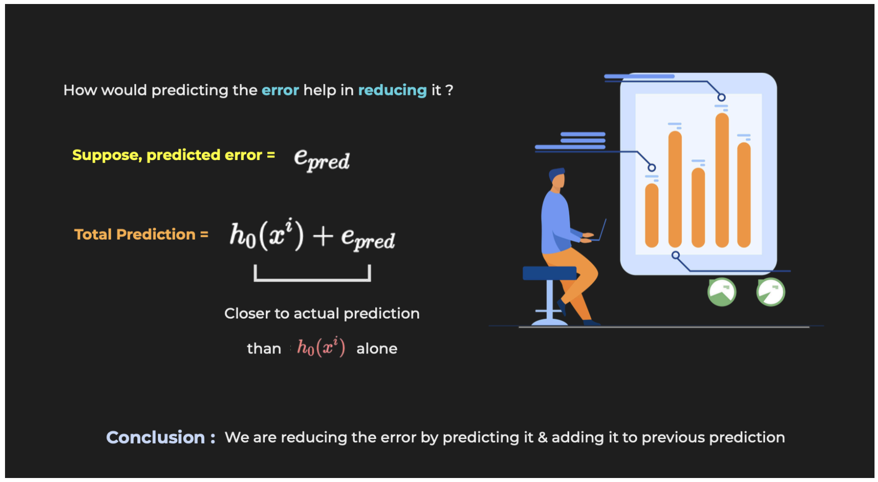
Process of creating a boosting ensemble

1. A simple mean model is used and trained on training data.

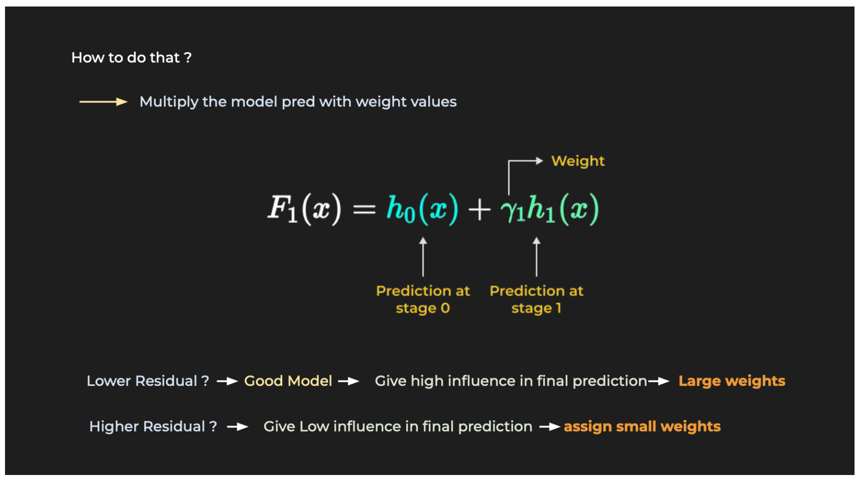


   - The error becomes $\rightarrow err_0^{(i)} = y^{(i)} - h_0(x^{(i)})$; where $i \in \{1,..N\}$
   - Therefore $y^{(i)} = h_0(x^{(i)}) + err_0^{(i)}$

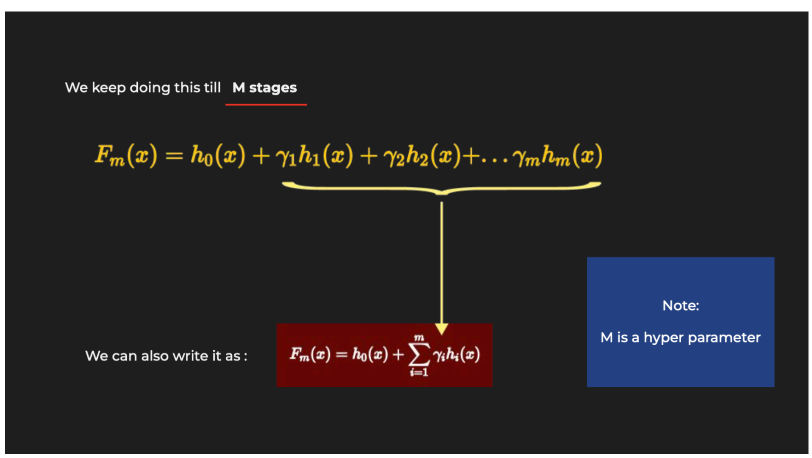
2. Now use a second model with the features x and target variable as the error from previous model err_0
   - Why make the error of the previous model as a target for the new model ?



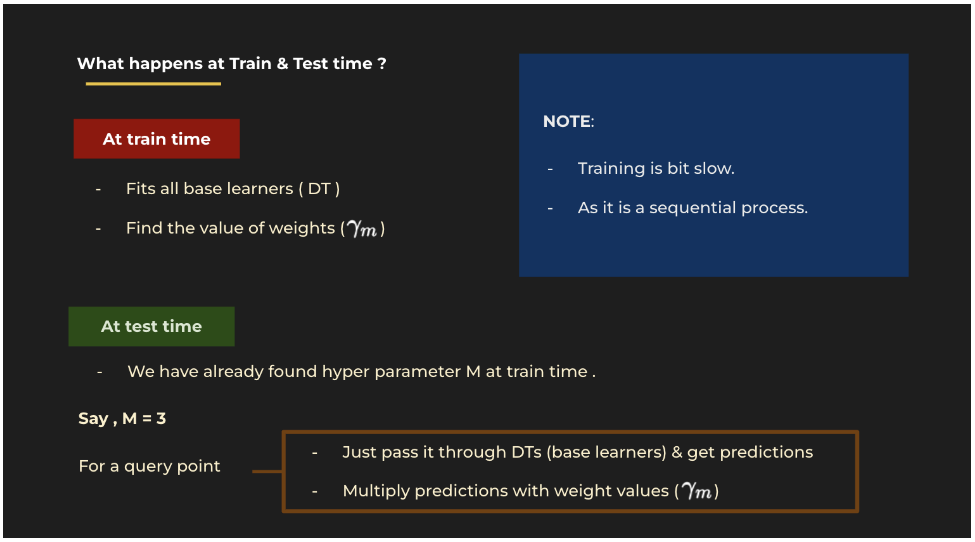
- With each addition of a new model a weight γ is associated



3. Repeat the addition of new models for M stages where M is an hyperparameter



Train/Test Time of Boosting

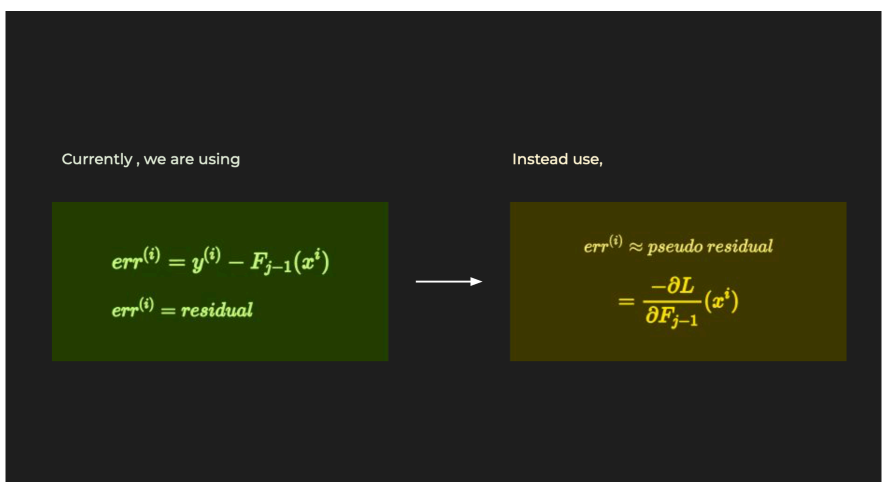


Gradient Boosting/Pseudo residual - Intuition

- As models are built sequentially rather than parallel, boosting uses gradient boosting algo to minimize the loss

How Gradient boosting algo reduces loss ?

- By using **Pseudo residuals** which is the negative gradient of the loss function with respect to the predicted values,



More content