Gradient Boosting

* Boosting algorithm → Regression & Classification (Ensemble)
* Core for XGBoost, Catboost, LGBM
* Jerome H Friedman (1999 - 2001)

General Working

Already discussed

Mathematical

X_1	X_2	y	Initial pred	Residuals	Weak learner prediction	Updated pred
1	2	5	7	-2	-1.05	6.85
2	1	6	7	-1	-1.05	6.85
3	4	7	7	0	1.05	7.15
4	3	10	7	3	1.05	7.15

① Initial Prediction

(Mean of target value)

$$\text{Mean} = \frac{5+6+7+10}{4} = 7.$$

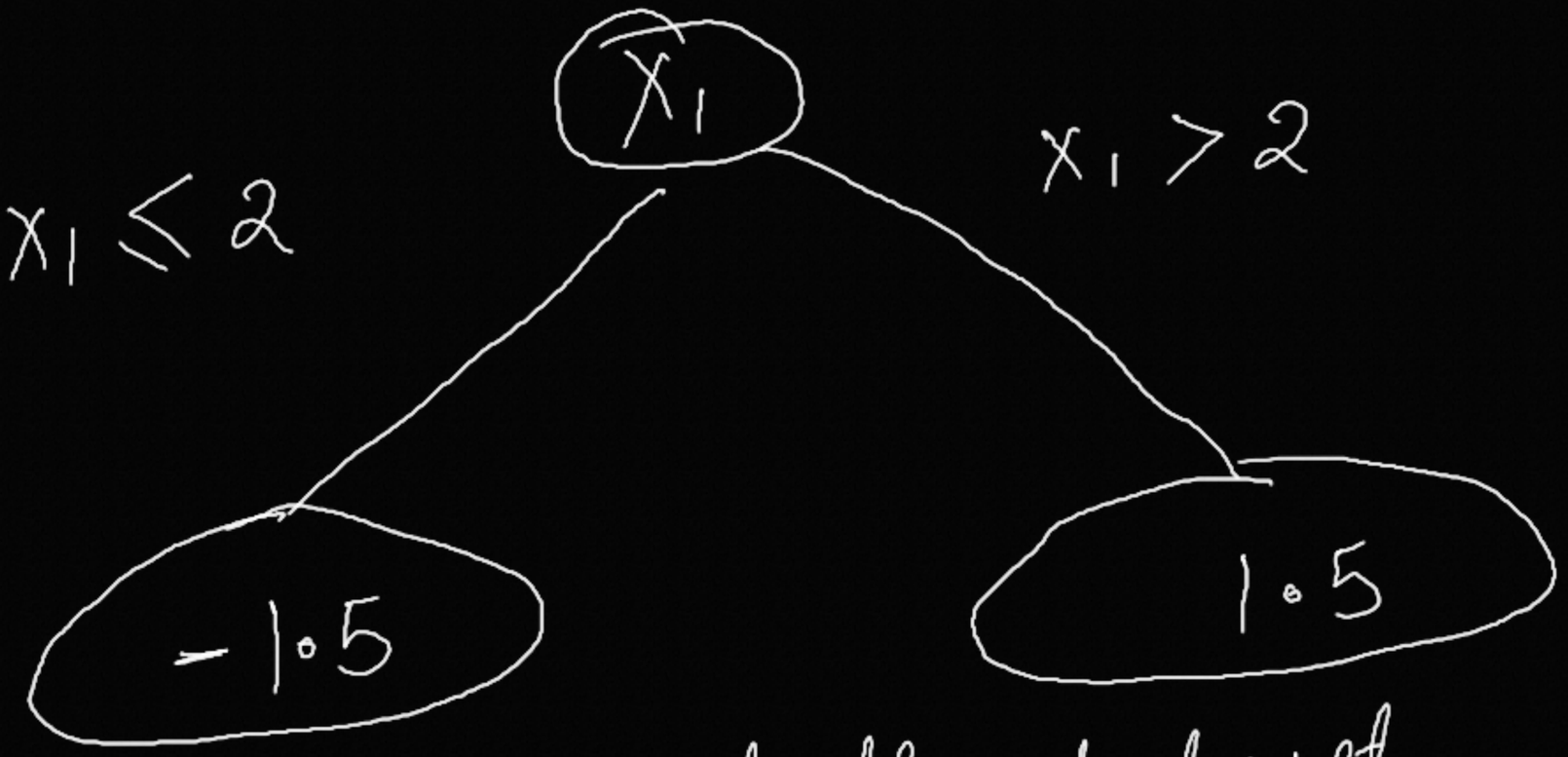
② Calculate residuals / errors

$$\text{Residual} = Y - \hat{Y}$$

(Actual value) (Predicted value)

③ Train weak learner (DT) to residuals

DT ⇒ Root Node → x_1 , Threshold = 2



Use model on full dataset

(4) Updated Predictions

Updated prediction = Initial pred + α × Weak learner prediction

$\alpha \rightarrow$ learning parameter (0.1)

1st row Updated pred = $7 + 0.1 \times (-1.5)$

= 6.85

⑤ Repeat Steps 2-4.

Until residual / error become very small.

Training Complete

Prediction → same as training

eg. $X_1 = 2.5$

$X_2 = 2.5$

Initial pred = 7

Apply weak learner ⇒ Pred ⇒ 1.5

updated prediction ⇒ $7 + 0.1 \times 1.5$

$= 7.15$

So on.

When to Use

* Complex & Non-linear datasets.
* Imbalanced datasets
  Missing data

Good performance with very less FE

When Not to Use

* Very large datasets
* Not good for realtime prediction (Slow response)
* Simple / linear data
* Require careful hyparameter tuning
* Outlier cant handle