

XGBoost Step-by-Step Example

Data

Given the following data:

Data Point (X)	True Value (Y)
1	3
2	5
3	7

Step 1: Initial Prediction

The initial prediction for all data points is the mean of the target values:

$$\hat{y}_i^{(0)} = \text{mean}(y) = \frac{3 + 5 + 7}{3} = 5$$

Thus, the initial predictions for all data points are:

$$\hat{y}_1^{(0)} = \hat{y}_2^{(0)} = \hat{y}_3^{(0)} = 5$$

Step 2: Compute Residuals (Errors)

Now, we compute the residuals (the difference between the true values and the initial predictions):

$$r_1^{(0)} = y_1 - \hat{y}_1^{(0)} = 3 - 5 = -2$$

$$r_2^{(0)} = y_2 - \hat{y}_2^{(0)} = 5 - 5 = 0$$

$$r_3^{(0)} = y_3 - \hat{y}_3^{(0)} = 7 - 5 = 2$$

Thus, the residuals are:

$$r_1^{(0)} = -2, \quad r_2^{(0)} = 0, \quad r_3^{(0)} = 2$$

Step 3: Fit a Tree to the Residuals

We now build a decision tree to predict the residuals. Let's assume the tree predicts the following values for the residuals:

$$f_1(x_1) = -1, \quad f_1(x_2) = 0, \quad f_1(x_3) = 1$$

Step 4: Update Predictions

We use a learning rate $\eta = 0.1$ to update the predictions. The new predictions are:

$$\hat{y}_i^{(1)} = \hat{y}_i^{(0)} + \eta f_1(x_i)$$

Thus, for each data point:

$$\hat{y}_1^{(1)} = 5 + 0.1 \times (-1) = 5 - 0.1 = 4.9$$

$$\hat{y}_2^{(1)} = 5 + 0.1 \times 0 = 5 + 0 = 5$$

$$\hat{y}_3^{(1)} = 5 + 0.1 \times 1 = 5 + 0.1 = 5.1$$

Thus, after updating with the first tree:

$$\hat{y}_1^{(1)} = 4.9, \quad \hat{y}_2^{(1)} = 5, \quad \hat{y}_3^{(1)} = 5.1$$

Step 5: Predict for New Data Point $X = 5$

Now, we predict for a new data point where $X = 5$. Since we have only one tree in this example, we will use the initial prediction and the tree's residual prediction for this new point.

Initial Prediction for $X = 5$

The initial prediction for any data point is the mean value, which we calculated earlier:

$$\hat{y}_5^{(0)} = 5$$

Residual Prediction for $X = 5$

We assume that the decision tree predicts a residual of 0.5 for this new point:

$$f_1(x_5) = 0.5$$

Update Prediction for $X = 5$

We update the prediction for $X = 5$ using the learning rate $\eta = 0.1$:

$$\hat{y}_5^{(1)} = \hat{y}_5^{(0)} + 0.1 \times 0.5 = 5 + 0.05 = 5.05$$

Final Prediction for $X = 5$

The prediction for the new data point $X = 5$ is:

$$\hat{y}_5^{(1)} = 5.05$$

Summary

- **Initial Prediction:** 5 for all data points.
- **Residuals (errors):** $r_1^{(0)} = -2, r_2^{(0)} = 0, r_3^{(0)} = 2$.
- **Tree Predictions** for residuals: $f_1(x_1) = -1, f_1(x_2) = 0, f_1(x_3) = 1$.
- **Updated Predictions** after first tree: $\hat{y}_1^{(1)} = 4.9, \hat{y}_2^{(1)} = 5, \hat{y}_3^{(1)} = 5.1$.
- **Prediction for New Data Point $X = 5$:** 5.05.