Gradient Boosting with EV Battery Life — Matrix Style

The Goal

Predict EV battery life — how many years it lasts — using features like:

- Temperature
- Usage
- Battery Age

We'll use matrix math to understand **Gradient Boosting**.

Example Data

Suppose we have data for 3 EVs and 2 features:

$$X = \begin{bmatrix} 25 & 2 \\ 30 & 3 \\ 20 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}$$

Columns in X: Temperature and Battery Age

Step 1: Initial Prediction

Start with the average battery life:

$$\hat{y}^{(0)} = \text{mean}(y) = \frac{5+4+6}{3} = 5$$

So, our initial prediction for all EVs:

$$\hat{y}^{(0)} = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix}$$

Step 2: Residuals (Errors)

Calculate how far off our predictions are:

$$r^{(1)} = y - \hat{y}^{(0)} = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Step 3: First Tree Learns Residuals

Tree $f_1(X)$ predicts:

$$f_1(X) = \begin{bmatrix} 0.1 \\ -0.9 \\ 0.8 \end{bmatrix}$$

These are close to the actual residuals.

Step 4: Update Predictions

Use learning rate $\eta = 0.1$:

$$\hat{y}^{(1)} = \hat{y}^{(0)} + \eta f_1(X) = \begin{bmatrix} 5 \\ 5 \\ 5 \end{bmatrix} + 0.1 \cdot \begin{bmatrix} 0.1 \\ -0.9 \\ 0.8 \end{bmatrix} = \begin{bmatrix} 5.01 \\ 4.91 \\ 5.08 \end{bmatrix}$$

Step 5: New Residuals

$$r^{(2)} = y - \hat{y}^{(1)} = \begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 5.01 \\ 4.91 \\ 5.08 \end{bmatrix} = \begin{bmatrix} -0.01 \\ -0.91 \\ 0.92 \end{bmatrix}$$

Step 6: Second Tree Learns New Residuals

Tree $f_2(X)$ predicts:

$$f_2(X) = \begin{bmatrix} -0.02 \\ -0.85 \\ 0.9 \end{bmatrix}$$

Step 7: Update Again

$$\hat{y}^{(2)} = \hat{y}^{(1)} + \eta f_2(X) = \begin{bmatrix} 5.01 \\ 4.91 \\ 5.08 \end{bmatrix} + 0.1 \cdot \begin{bmatrix} -0.02 \\ -0.85 \\ 0.9 \end{bmatrix} = \begin{bmatrix} 5.008 \\ 4.825 \\ 5.17 \end{bmatrix}$$

Repeat!

Each round:

- Compute new residuals
- Train tree to fix them
- Update predictions slowly

This brings predictions closer to the true y.

Final Summary

$$X \in \mathbb{R}^{n \times d}, \quad y \in \mathbb{R}^n$$

$$\hat{y}^{(0)} = \text{mean}(y) \cdot \mathbf{1}$$

$$r^{(t)} = y - \hat{y}^{(t-1)}, \quad f_t(X) \approx r^{(t)}$$

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(X)$$

This is how Gradient Boosting works — step by step!