Gradient Boosting for Regression: BMI Prediction Example

Gradient Boosting for Regression: BMI Prediction Example

Let's use a synthetic dataset to predict **BMI (Body Mass Index)** using **height (m)** and **weight (kg)**. BMI is calculated as $\text{BMI} = \frac{\text{weight}}{\text{height}^2}$. We'll apply gradient boosting with **Mean Squared Error (MSE)** loss and a **learning rate** of $\gamma = 0.1$.

Dataset

Height (m)	Weight (kg)	BMI (Target)
1.6	50	19.53
1.7	60	20.76
1.8	70	21.60
1.9	80	22.16

Goal: Predict BMI using gradient boosting (2 iterations, tree depth = 1).

Step 1: Initial Model (F_0)

- Start with the **mean BMI**:

$$F_0(x) = \frac{19.53 + 20.76 + 21.60 + 22.16}{4} \approx 21.01$$

- Initial prediction for all samples: **21.01**.

Residuals (r = Actual BMI $- F_0(x)$):

$$\text{Residuals} = [-1.48, -0.25, 0.59, 1.15]$$

Step 2: First Tree (h_1)

Fit a regression tree (depth = 1) to predict residuals using **height** and **weight**.

Possible Splits:

1. **Split on Weight 65 kg**:
   - Left node (Weight 65): Samples 1 & 2. Average residual $= \frac{-1.48+(-0.25)}{2} \approx -0.865$.
   - Right node (Weight ¿ 65): Samples 3 & 4. Average residual $= \frac{0.59+1.15}{2} \approx 0.87$.
2. **Split on Height 1.75 m**:
   - Left node (Height 1.75): Samples 1 & 2. Average residual $= -0.865$.
   - Right node (Height ¿ 1.75): Samples 3 & 4. Average residual $= 0.87$.

Choose the split with the lowest MSE. Both splits give the same residuals, so we pick **Weight 65 kg**.

Tree $h_1(x)$ Predictions:

- If Weight 65 kg: Predict -0.865.
- If Weight ¿ 65 kg: Predict 0.87.

Step 3: Update Model (F_1)

$$F_1(x) = F_0(x) + \gamma \cdot h_1(x)$$

- For Weight 65 kg (Samples 1 & 2): $21.01 + 0.1 \cdot (-0.865) \approx 20.92$.
- For Weight ¿ 65 kg (Samples 3 & 4): $21.01 + 0.1 \cdot 0.87 \approx 21.10$.

New Predictions:

$$\text{Predicted BMI} = [20.92, 20.92, 21.10, 21.10]$$

New Residuals ($r = \text{Actual BMI} - F_1(x)$):

$$\text{Residuals} = [-1.39, -0.16, 0.50, 1.06]$$

Step 4: Second Tree (h_2)

Fit another tree to the updated residuals.

Possible Splits:

1. **Split on Height 1.75 m**:
   - Left node (Height 1.75): Samples 1 & 2. Average residual $= \frac{-1.39+(-0.16)}{2} \approx -0.775$.
   - Right node (Height ¿ 1.75): Samples 3 & 4. Average residual $= \frac{0.50+1.06}{2} \approx 0.78$.
2. **Split on Weight 75 kg**:
   - Left node (Weight 75): Samples 1, 2, 3. Average residual $= \frac{-1.39+(-0.16)+0.50}{3} \approx -0.35$.
   - Right node (Weight ¿ 75): Sample 4. Residual = 1.06.

Best split: **Height 1.75 m** (lower MSE).

Tree $h_2(x)$ Predictions:

- If Height 1.75 m: Predict -0.775.
- If Height ¿ 1.75 m: Predict 0.78.

Step 5: Update Model (F_2)

$$F_2(x) = F_1(x) + \gamma \cdot h_2(x)$$

- For Height 1.75 m (Samples 1 & 2): $20.92 + 0.1 \cdot (-0.775) \approx 20.84$.
- For Height ¿ 1.75 m (Samples 3 & 4): $21.10 + 0.1 \cdot 0.78 \approx 21.18$.

Final Predictions:

$$\text{Predicted BMI} = [20.84, 20.84, 21.18, 21.18]$$

Final Residuals:

$$\text{Residuals} = [-1.31, -0.08, 0.42, 0.98]$$

Key Observations

- **Residuals Shrink**: Errors decrease with each iteration (e.g., Sample 1's residual went from -1.48 to -1.31).
- **Feature Importance**: The model prioritized **weight** first, then **height**, reflecting their correlation with BMI.
- **Learning Rate (γ)**: A small $\gamma = 0.1$ ensures gradual corrections, preventing overfitting.

How Gradient Boosting Works

- **Sequential Correction**: Each tree fixes the errors of the previous ensemble.
- **Gradient Descent**: Residuals are the negative gradients of the MSE loss function.
- **Additive Model**: Predictions are refined incrementally:

$$F_M(x) = F_0(x) + \gamma \sum_{m=1}^{M} h_m(x)$$

Gradient boosting builds an ensemble of weak trees to minimize residuals iteratively. In this example, after two iterations, predictions improved by incorporating patterns in weight and height. With more trees and deeper splits, the model would further reduce residuals, approaching the true BMI values.