

REGRESSION TREES

Differences from Classification Trees

- **Target:** Continuous values (e.g., temperature, price).
- **Splitting Criterion:** Minimize **MSE** or **MAE**.

Example: Predicting House Prices

- **Split 1:** If Sq. Ft > 1500 → Predict \$350k.
- **Split 2:** Else if Location = "Urban" → Predict \$275k.

Mathematical Formulation

- **MSE Calculation:**

$$\text{MSE} = \frac{1}{N} \sum (y_i - \hat{y})^2, \text{ where } \hat{y} \text{ is the node's mean value.}$$

Mathematical Formulation

- **MSE Calculation:**

$$\text{MSE} = \frac{1}{N} \sum (y_i - \hat{y})^2$$
, where $\hat{y}$ is the node's mean value.

Advantages

- Handles non-linear relationships.
- Robust to outliers (vs. linear regression).

Limitations

- Poor extrapolation beyond training range.