Regression Trees – Decision Trees for Continuous Outputs

---

1. Introduxtion

- **Goal:** Predict *continuous numerical values* (like house prices, temperature, sales).
- Unlike classification trees (predict categories), **Regression Trees** output **a number** at each leaf.
- The tree structure partitions the data into regions, and each region has a constant prediction (mean value).

---

2. Best Fit Line

- In linear regression, we fit a **line** to minimize **Mean Squared Error (MSE)** between predictions and actual values.
- **Regression Trees** take a different approach:
  1. Instead of a single line, they split the data into segments.
  2. In each segment, the prediction is just the **average of the target values** in that segment.

---

3. Finding the Splitting Criteria

- **Key idea:** Choose splits that make the data in each side as *homogeneous* as possible in terms of target values.
- **How we measure it:**
  - ❖ Use **Variance Reduction** or **MSE Reduction**.
  - ❖ At each step:
    1. Try all possible features and split points.
    2. For each split, calculate MSE for left and right groups.
    3. Pick the split that **reduces total MSE the most**.

---

4. Final Criteria

- The **best split** is the one that **maximizes the reduction in prediction error** (MSE drop).
- The splitting process stops when:
  1. No further MSE reduction possible, OR
  2. Node has too few samples (controlled by hyperparameters like min_samples_split).

---

5. Streamlit Web App

- Visual demo where you can input feature values and see the regression tree's prediction.
- Helps understand **decision boundaries** in regression tasks — boundaries are **flat steps**, not sloped like linear regression.

---

6. Code Example

- Shows implementation in libraries like sklearn.tree.DecisionTreeRegressor.
- Steps:
    1. Import and initialize model.
    2. Fit on training data.
    3. Predict on test data.
    4. Evaluate using metrics like **MSE**, **RMSE**, or **R^2 score**.

---

7. Feature Importance

- Regression Trees can rank features based on their contribution to reducing prediction error.
- Feature importance is calculated from the **total MSE reduction** each feature provides across all splits.
- Useful for **interpretability** and **feature selection**.

➢ Key Takeaway

- **Regression Trees** don't fit a single function; they split the data space into constant-valued regions.
- They work well for capturing **non-linear relationships**.
- They're **easy to interpret**, but without control (depth, min samples), they can overfit just like classification trees.
