Decision Tree Regression

Definition - Decision Tree Regression is a machine learning algorithm used to predict continuous values by recursively splitting the data into subsets based on feature values. Unlike linear models, it captures non-linear relationships between features and the target variable by creating decision nodes and leaves.

Feature Scaling - Feature scaling is **not required** in Decision Tree Regression because it is not a distance-based algorithm. The splits in a decision tree are based on comparisons of feature values to thresholds, making the raw scale of the features irrelevant.

How It Works

- At the first step, it looks at all available features and finds the one that best divides the data to minimize the error in the target variable. For regression, this is typically done by selecting the feature that results in the largest reduction in Mean Squared Error (MSE). The selected feature becomes a **decision node**.
- The data is then split based on a threshold of the selected feature. For example, if the selected feature is "square footage," the data might be split into two groups: one with houses greater than 1500 sq ft and one with houses less than or equal to 1500 sq ft.
- For each subset of data resulting from the split, the process repeats. The algorithm evaluates the remaining features and selects the next best feature to split on, further refining the partitions in the data.
- Eventually, the recursion reaches a stopping point. This could be when all data points in a subset have the same target value, a pre-defined maximum depth is reached, or further splits no longer improve the prediction. At this point, the subset becomes a **leaf node**, where the target value is predicted by averaging the outcomes within that subset.
- Once all splits are made, the tree structure is complete, with decision nodes representing feature splits and leaf nodes representing predicted values.
- When making predictions for new data, the model starts at the root of the tree and follows the decision nodes, applying the same thresholds to the new input data until it reaches a leaf node. The value at the leaf node becomes the prediction for the target variable.

In this way, Decision Tree Regression builds a model that breaks down complex data relationships into a series of if-else conditions that lead to a continuous value prediction.