



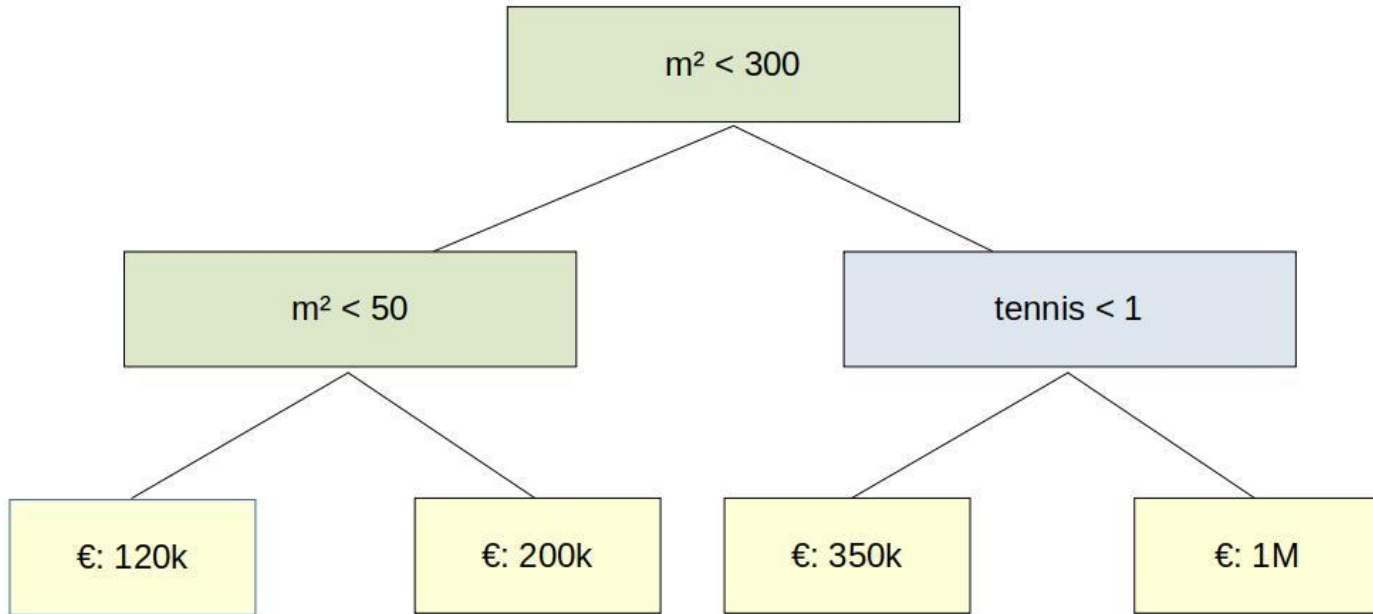
Tree-Based Methods

Decision Tree: Dataset



m ²	Tennis	Sale Price
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

Decision Tree (DT)



Building a Decision Tree



- Iterate over each feature X_i
 - Iterate over all possible values V_i
 - Split the data based on V_i threshold (less than V_i and greater than V_i).
 - Each one of the 2 splits predicts the average of their elements.
 - Calculate metric (e.g., MSE) for each of the splits \rightarrow MSE_L, MSE_R.
 - Calculate final metric MSE_total.
- Use the split with less MSE_total.



Decision Tree: MSE_total

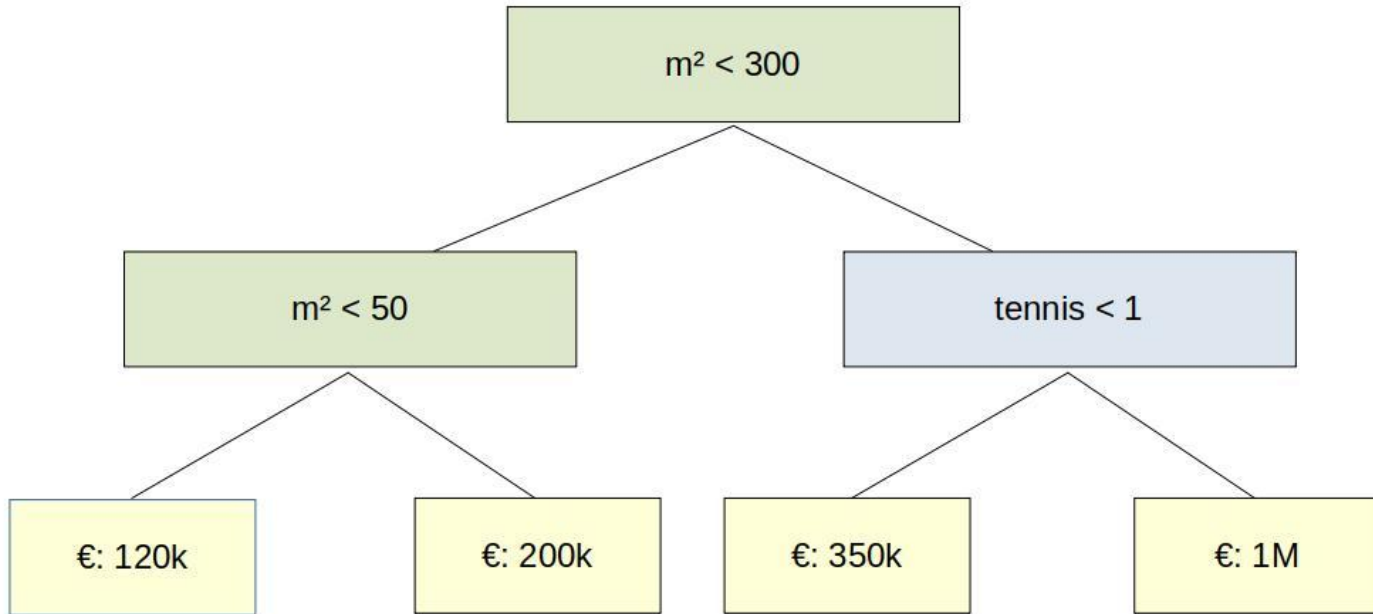
$$\text{MSE}_{\text{total}} = \frac{\text{size}(\text{Left})}{\text{size}(\text{Left}) + \text{size}(\text{Right})} \times \text{MSE}_L + \frac{\text{size}(\text{Right})}{\text{size}(\text{Left}) + \text{size}(\text{Right})} \times \text{MSE}_R$$

Decision Tree: Dataset



m ²	Tennis	Sale Price
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

Decision Tree (DT)

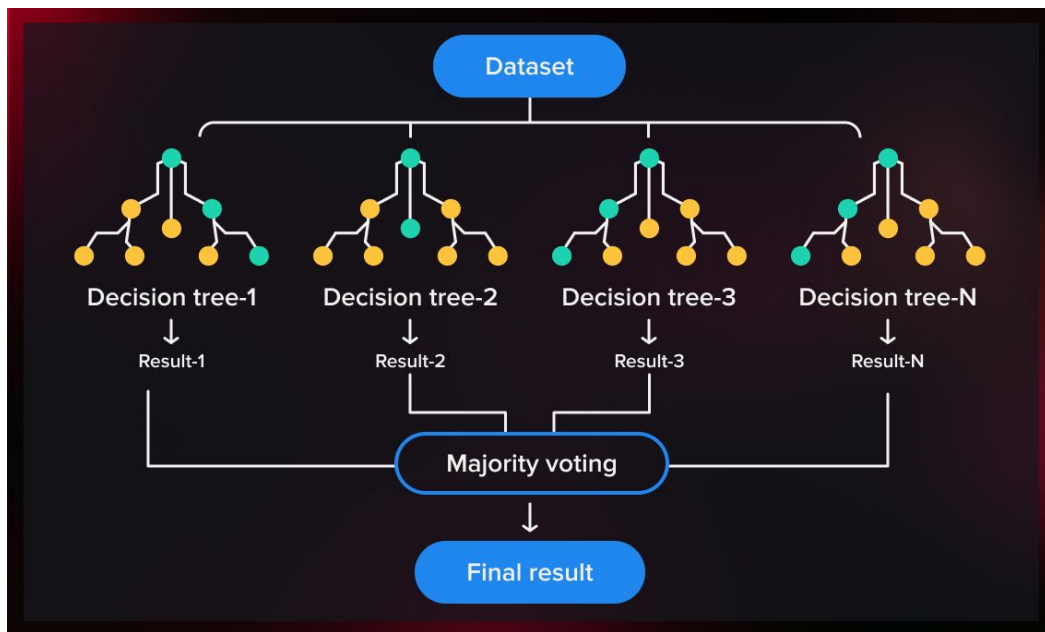




DT: Important Hyperparams.

- Criterion (metric): squared error, absolute error, etc.
- Max. tree depth.
- Min. samples for splitting.
- Min. samples per leaf.
- Max. leaf nodes.

Random Forest (RF)

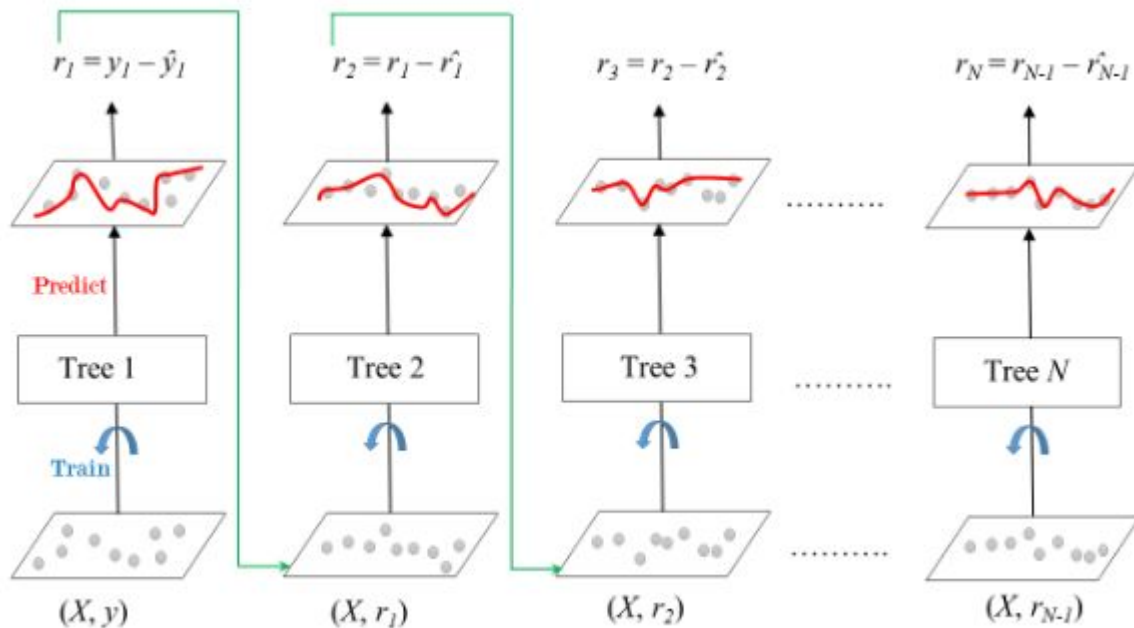




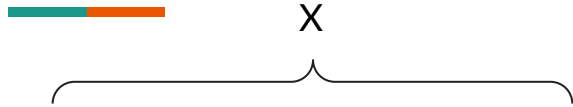
RF: Important Hyperparams.

- Num. estimators.
- Criterion (metric): squared error, absolute error, etc.
- Max. tree depth.
- Min. samples for splitting.
- Min. samples per leaf.
- Max. features.
- Max. samples.

Gradient Boosted Decision Trees (GBDT)



GBDT: Dataset



A diagram consisting of a horizontal bracket with a small upward-pointing hook at its center. Above the bracket is the letter 'X'. The bracket spans the width of the first two columns of the table below, indicating that these two columns represent the feature set X.

m ²	Tennis	Sale Price (y)
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

GBDT: Dataset

X			1st prediction
m ²	Tennis	Sale Price (y)	$\bar{y}_1 = F_0$
50	0	130k	329k
65	0	250k	329k
25	0	70K	329k
850	1	1M	329k
55	0	150	329k
290	0	300k	329k
430	0	400k	329k

GBDT: Dataset


X			1st residuals	
m ²	Tennis	Sale Price (y)	$\hat{y}_1 = F_0$	$r_1 = y - \hat{y}_1$
50	0	130k	329k	199k
65	0	250k	329k	79k
25	0	70K	329k	259k
850	1	1M	329k	-671k
55	0	150	329k	179k
290	0	300k	329k	29k
430	0	400k	329k	-71k

GBDT: Dataset



m^2	Tennis	Sale Price (y)
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

Train tree with X and r1:



$\hat{y}_1 = F_0$	$r_1 = y - \hat{y}_1$
329k	199k
329k	79k
329k	259k
329k	-671k
329k	179k
329k	29k
329k	-71k

GBDT: Dataset



X



m ²	Tennis	Sale Price (y)
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

New prediction = $y_2 = F_0 + h(X, r_1)$

$\bar{y}_1 = F_0$	$r_1 = y - \bar{y}_1$
329k	199k
329k	79k
329k	259k
329k	-671k
329k	179k
329k	29k
329k	-71k



GBDT: Prediction Equation

$$\hat{F}_M(X) = \hat{y} = F_0(X) + \alpha \sum_{m=1}^M h_m(X, r_m)$$

Prediction

Prediction

Learning rate

Aggregates the contributions of all trees



GBDT: Important Hyperparams.

- Num. estimators.
- Criterion (metric): squared error, absolute error, etc.
- Learning rate.
- Max. tree depth.
- Min. samples for splitting.
- Min. samples per leaf.
- Max. features.
- Subsample.