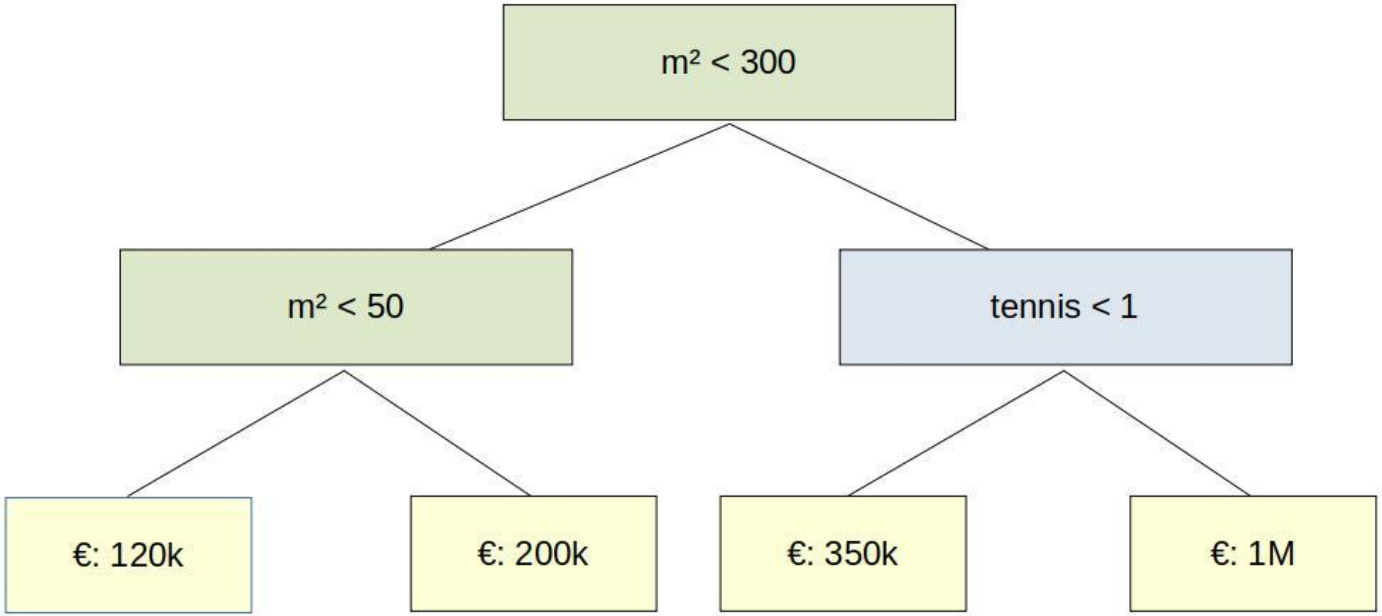
Tree-Based Methods

Decision Tree: Dataset

m²	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 Xi
  - Iterate over all possible values Vi
    - Split the data based on Vi threshold (less than Vi and greater than Vi).
    - Each one of the 2 splits predicts the average of their elements.
    - Calculate metric (e.g., MSE) for each of the splits → MSE_L, MSE_R.
    - Calculate final metric MSE_total.
- Use the split with less MSE_total.

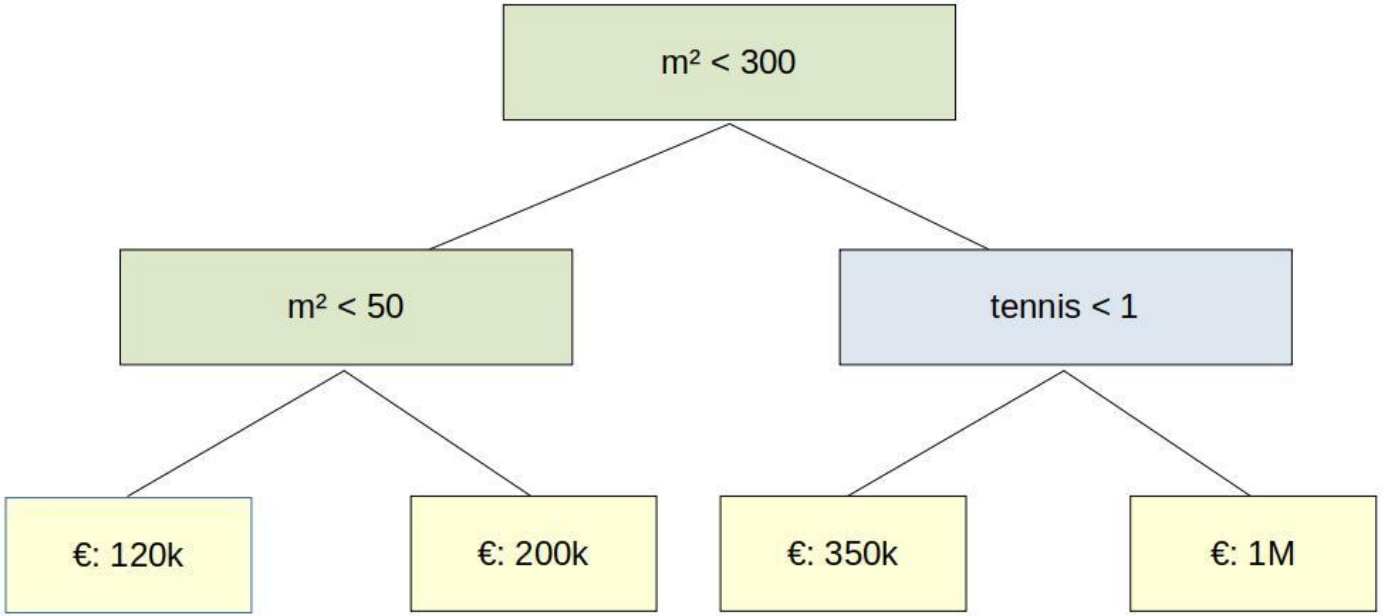
Decision Tree: MSE_total

$$\mathrm{MSE}_{\mathrm{total}} = \frac{\mathrm{size}(\mathrm{Left})}{\mathrm{size}(\mathrm{Left}) + \mathrm{size}(\mathrm{Right})} \times \mathrm{MSE}_{\mathrm{L}} + \frac{\mathrm{size}(\mathrm{Right})}{\mathrm{size}(\mathrm{Left}) + \mathrm{size}(\mathrm{Right})} \times \mathrm{MSE}_{\mathrm{R}}$$

Decision Tree: Dataset

m²	Tennis	Sale Price
50	0	130k
65	0	250k
25	0	70K
850	1	1M
55	0	150
290	0	300k
430	0	400k

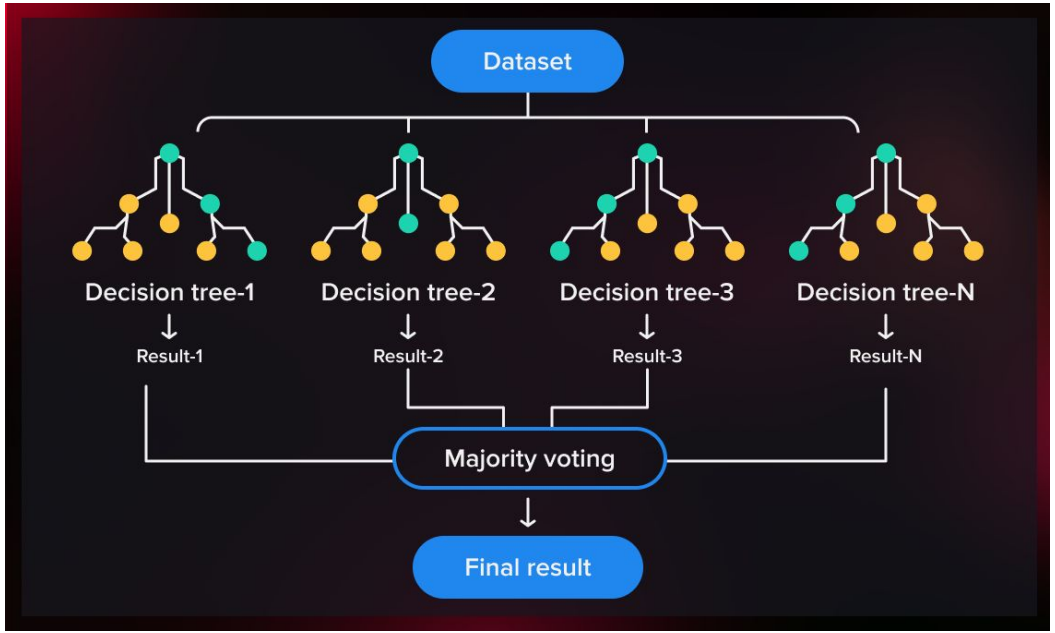
Decision Tree (DT)



DT: Important Hyperparams.

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

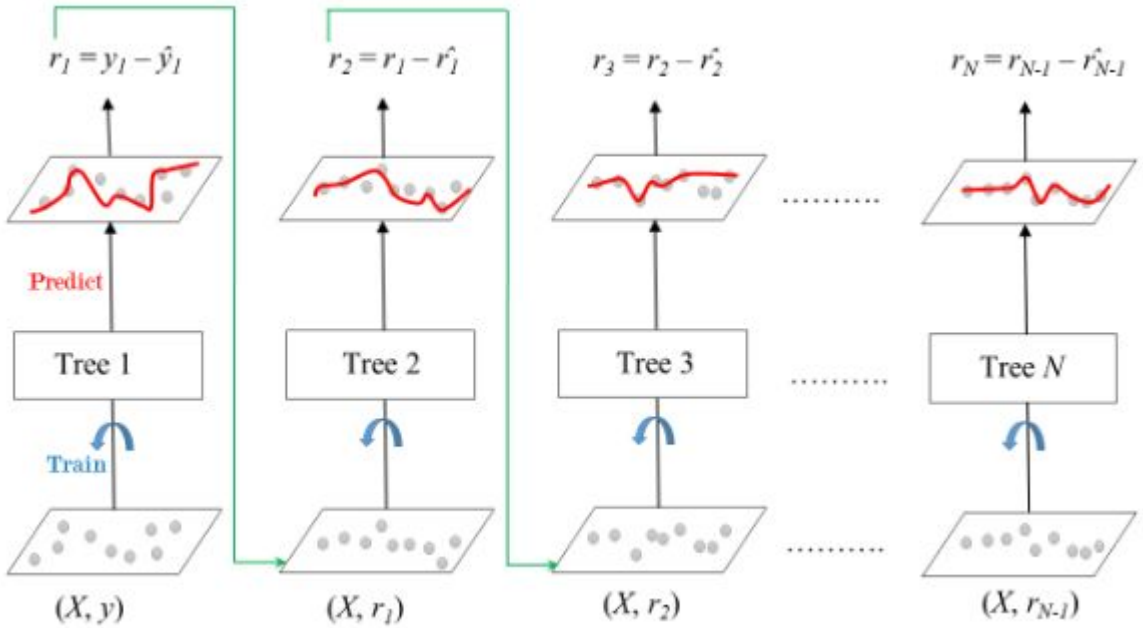
Random Forest (RF)



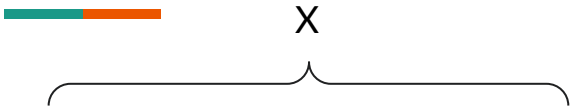
RF: Important Hyperparams.

- 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



m²	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

m²	Tennis	Sale Price (y)	1st prediction ȳ1 = F0
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

m²	Tennis	Sale Price (y)	ȳ1 = F0	r1 = y - ȳ1
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

X

Train tree with X and r1:
h1(x)

m²	Tennis	Sale Price (y)	ȳ1 = F0	r1 = y - ȳ1
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

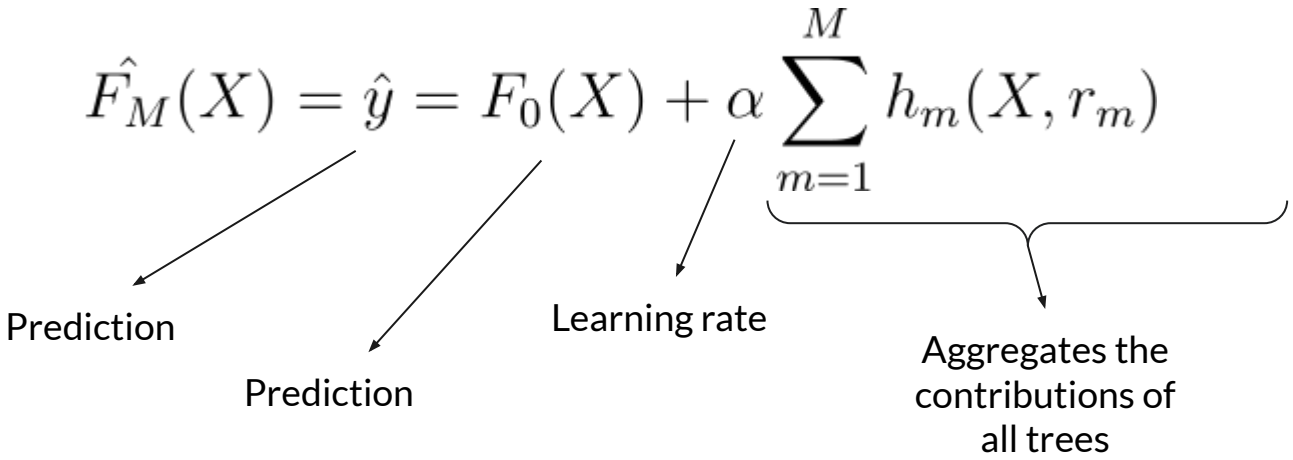
GBDT: Dataset

X

New prediction = y2 = F0 + h(X, r1)

m²	Tennis	Sale Price (y)	ȳ1 = F0	r1 = y - ȳ1
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: Prediction Equation



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.