



Data Analysis Block Course: Statistics and Machine learning

Decision Trees by Hand - Solutions

Prof. Dr. Mikhail Mikhaseenko

Assistants: Hendric Jonas, Christoph Tönnis



28 September 2026

Indico schedule

RUB Moodle

Part I - Understanding a single leaf

Solution 1.1 - One constant prediction

For the targets 2, 3, 4, 7, the constant prediction and errors are

$$\begin{aligned}\bar{y} &= (2 + 3 + 4 + 7)/4 = 4 \\ (r_1, r_2, r_3, r_4) &= (-2, -1, 0, 3) \\ \text{SSE} &= 4 + 1 + 0 + 9 = 14 \\ \text{MSE} &= 14/4 = 3.5\end{aligned}$$

The MSE is the average squared deviation, in squared target units.

Solution 1.2 - Is the mean special?

For a non-empty leaf, differentiate the squared-error loss and solve the stationary-point equation:

$$\begin{aligned}L(c) &= \sum_{i=1}^N (y_i - c)^2 \\ L'(c) &= -2 \sum_{i=1}^N (y_i - c) = 2Nc - 2 \sum_{i=1}^N y_i \\ L'(c) = 0 &\implies c = \frac{1}{N} \sum_{i=1}^N y_i \\ L''(c) &= 2N > 0\end{aligned}$$

Thus the mean is the unique global minimizer. Dividing the loss by N leaves the minimizer unchanged.

Solution 1.3 - What does MSE actually measure?

1. MSE has squared target units.
2. Doubling a residual multiplies its squared-error contribution by four.
3. Every residual is zero, so all targets equal their predictions.
4. $\text{RMSE} = \sqrt{3.5} \approx 1.87$, in the same units as the target.

Solution 1.4 - Could we use another criterion?

1. The targets are 2, 3, 4, 7. Any value in $[3, 4]$ minimizes the absolute error. Their conventional median is $(3 + 4)/2 = 3.5$.
2. The MAE is

$$\text{MAE}(3.5) = \frac{1.5 + 0.5 + 0.5 + 3.5}{4} = 1.5$$

Part II - How does a regression tree find its first split?

Solution 2.1 - Generate the training data by hand

The training data are

x	$y = x^2$
0	0
1	1
2	4
3	9
4	16
5	25
6	36

The root prediction is

$$\bar{y} = \frac{0 + 1 + 4 + 9 + 16 + 25 + 36}{7} = 13$$

The plots in Solution 2.9 show the training points and the true parabola.

Solution 2.2 - Before splitting: the root leaf

The root SSE is

$$\text{SSE}_{\text{root}} = 1092$$

The root leaf predicts 13 everywhere, giving the horizontal line $\hat{f}(x) = 13$. The root MSE is $1092/7 = 156$.

Solution 2.3 - Which split points does the algorithm need to test?

The relevant candidate thresholds are the midpoints between neighboring observations:

$$0.5, 1.5, 2.5, 3.5, 4.5, 5.5$$

Any two thresholds between the same neighboring observations produce exactly the same assignment of training points, which is why only one representative threshold is needed.

Solution 2.4 - Follow one candidate split step by step

For $t = 2.5$:

Left child:

$x < 2.5$	y
0	0
1	1
2	4

$$\hat{y}_L = \frac{5}{3}$$

Right child:

$x \geq 2.5$	y
3	9
4	16
5	25
6	36

with prediction

$$\hat{y}_R = 21.5$$

The errors are

$$\text{SSE}_L = 26/3 \approx 8.67$$

$$\text{SSE}_R = 409$$

$$\text{SSE}_{\text{split}} = 1253/3 \approx 417.67$$

The reduction is $1092 - 1253/3 = 2023/3 \approx 674.33$.

Solution 2.5 - Let the tree choose the best split

The full comparison is

threshold t	mean left	mean right	SSE_L	SSE_R	total SSE
0.5	0.00	15.17	0.00	894.83	894.83
1.5	0.50	18.00	0.50	654.00	654.50
2.5	1.67	21.50	8.67	409.00	417.67
3.5	3.50	25.67	49.00	200.67	249.67
4.5	6.00	30.50	174.00	60.50	234.50
5.5	9.17	36.00	474.83	0.00	474.83

Therefore the best root split is

$$x < 4.5$$

The resulting depth-1 predictor is

$$\hat{f}(x) = \begin{cases} 6 & x < 4.5 \\ 30.5 & x \geq 4.5 \end{cases}$$

This is already a step function.

The error reduction is $1092 - 234.5 = 857.5$. The root tests $x < 4.5$: its left leaf predicts 6 and its right leaf predicts 30.5.

Solution 2.6 - Split the left child

After the first split, the left child contains

$$x = (0, 1, 2, 3, 4) \quad y = (0, 1, 4, 9, 16)$$

Its candidate thresholds are

$$0.5, 1.5, 2.5, 3.5$$

The corresponding total SSE values are approximately

threshold	total SSE
0.5	129.00
1.5	73.17
2.5	33.17
3.5	49.00

Thus the left child selects

$$x < 2.5$$

Solution 2.7 - Split the right child

The right child contains only

$$x = (5, 6) \quad y = (25, 36)$$

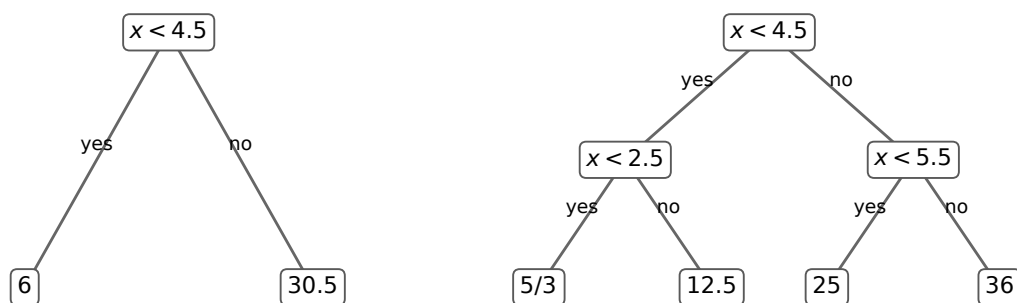
There is only one candidate threshold:

$$x < 5.5$$

After that split each of these two observations occupies its own leaf, so the SSE of this branch becomes zero.

Solution 2.8 - Draw the depth-2 tree

The depth-2 tree is therefore



The leaf predictions come from

$$\frac{0 + 1 + 4}{3} \approx 1.67$$

$$\frac{9 + 16}{2} = 12.5$$

and the single leaves 25 and 36.

The total training SSE is $199/6 \approx 33.17$, compared with 234.5 at depth 1. The depth-2 prediction is

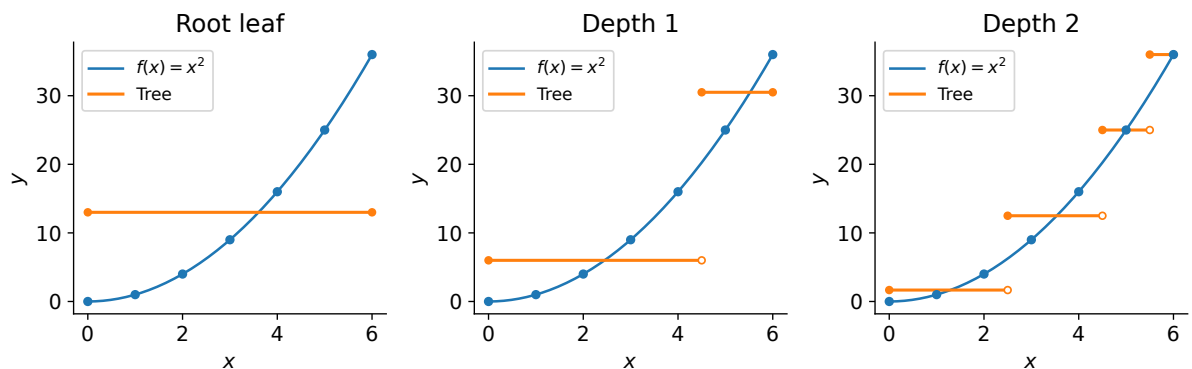
$$\hat{f}(x) = \begin{cases} 5/3 & x < 2.5 \\ 12.5 & 2.5 \leq x < 4.5 \\ 25 & 4.5 \leq x < 5.5 \\ 36 & x \geq 5.5 \end{cases}$$

Draw these horizontal segments over $0 \leq x \leq 6$, using the right-hand region's value at each threshold. The exact total SSE is $199/6$.

Solution 2.9 - Compare the approximations

Model	Training SSE
One root leaf	1092
Depth-1 tree	234.5
Depth-2 tree	$199/6 \approx 33.17$

The depth-2 tree has the smallest training error.



Part III - More than one input feature

Solution 3.1 - Understand the function before building a tree

For

$$f(x_1, x_2) = x_1^2 + 4x_2$$

the data are

x_1	x_2	y
0	0	0
1	0	1
2	0	4
3	0	9
0	1	4
1	1	5
2	1	8
3	1	13

For $x_2 = 0$,

$$f(x_1, 0) = x_1^2$$

while for $x_2 = 1$,

$$f(x_1, 1) = x_1^2 + 4$$

Thus the second curve is shifted upward by 4.

Solution 3.2 - Which feature is chosen first?

The root candidate thresholds are

$$x_1 < 0.5 \quad x_1 < 1.5 \quad x_1 < 2.5$$

and

$$x_2 < 0.5$$

The split comparison is

feature	threshold	mean left	mean right	total SSE
x_1	0.5	2.00	6.67	97.33
x_1	1.5	2.50	8.50	58.00
x_1	2.5	3.67	11.00	49.33
x_2	0.5	3.50	7.50	98.00

Therefore the root selects

$$x_1 < 2.5$$

The depth-1 tree predicts the same value for $x_2 = 0$ and $x_2 = 1$ whenever the two observations fall into the same x_1 region, because the root split has not used x_2 yet.

Solution 3.3 - Grow one more level

The left child contains (x_1, x_2) with $x_1 \in \{0, 1, 2\}$ and $x_2 \in \{0, 1\}$, with targets 0, 1, 4, 4, 5, 8. Its candidate comparison is:

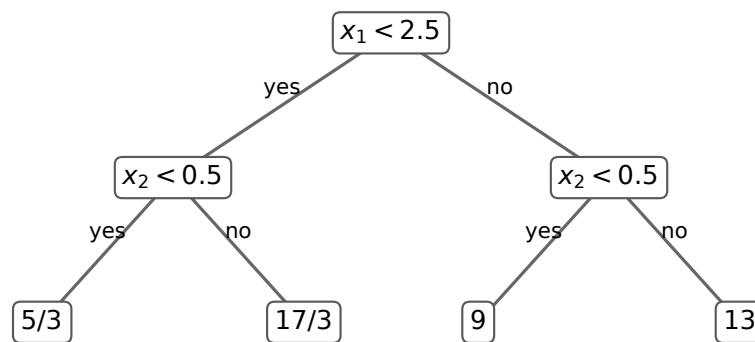
Feature	Threshold	Mean left	Mean right	Total SSE
x_1	0.5	2.00	4.50	33.00
x_1	1.5	2.50	6.00	25.00
x_2	0.5	1.67	5.67	17.33

The best split is $x_2 < 0.5$. Its leaves predict $5/3$ and $17/3$, and their total SSE is $52/3$.

The right child contains $(3, 0, 9)$ and $(3, 1, 13)$. The feature x_1 is constant here, so it cannot split the node into two non-empty children:

Feature	Threshold	Mean left	Mean right	Total SSE
x_2	0.5	9.00	13.00	0.00

The only candidate is $x_2 < 0.5$; it gives leaf predictions 9 and 13 and zero SSE.

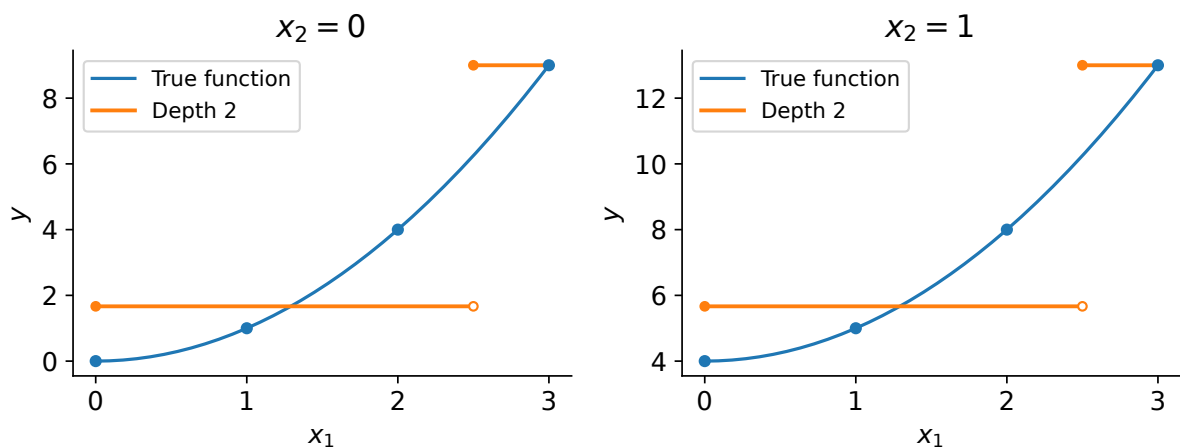


Every left branch means the condition is true; every right branch means it is false. For the requested sketches,

$$\hat{f}(x_1, 0) = \begin{cases} 5/3 & x_1 < 2.5 \\ 9 & x_1 \geq 2.5 \end{cases} \quad \hat{f}(x_1, 1) = \begin{cases} 17/3 & x_1 < 2.5 \\ 13 & x_1 \geq 2.5 \end{cases}$$

The two step functions now differ by 4, reflecting the effect of the second feature. Draw the curves x_1^2 and $x_1^2 + 4$ behind them for comparison.

The total training SSE falls from $148/3 \approx 49.33$ at depth 1 to $52/3 \approx 17.33$ at depth 2. Each individual split uses one feature, while different nodes can use different features. Together these splits define rectangular regions, with a constant prediction in each region.



Part IV - A forest by hand

Solution 4.1 - Training data

x	$f(x)$
0	0
$\pi/2$	$\frac{3}{2}$
π	1
$3\pi/2$	$\frac{1}{2}$
2π	2

There are five observations with no repetitions. For the sketch, add the sine curve to the line $y = x/\pi$.

Solution 4.2 - Train a tree using absolute error

The completed tables give the target values in increasing input order. Sort them within each child to determine its median.

Threshold $t = \pi/4$

Child		1	2	3	4	Median	MAE
L	x_i	0	---	---	---	---	---
	$f(x_i)$	0	---	---	---	0	0
R	x_i	$\pi/2$	π	$3\pi/2$	2π	---	---
	$f(x_i)$	$\frac{3}{2}$	1	$\frac{1}{2}$	2	$\frac{5}{4}$	$\frac{1}{2}$

Threshold $t = 3\pi/4$

Child		1	2	3		Median	MAE
L	x_i	0	$\pi/2$	---		---	---
	$f(x_i)$	0	$\frac{3}{2}$	---		$\frac{3}{4}$	$\frac{3}{4}$
R	x_i	π	$3\pi/2$	2π		---	---
	$f(x_i)$	1	$\frac{1}{2}$	2		1	$\frac{1}{2}$

Threshold $t = 5\pi/4$

Child		1	2	3		Median	MAE
L	x_i	0	$\pi/2$	π		---	---
	$f(x_i)$	0	$\frac{3}{2}$	1		1	$\frac{1}{2}$
R	x_i	$3\pi/2$	2π	---		---	---
	$f(x_i)$	$\frac{1}{2}$	2	---		$\frac{5}{4}$	$\frac{3}{4}$

Threshold $t = 7\pi/4$

Child		1	2	3	4	Median	MAE
L	x_i	0	$\pi/2$	π	$3\pi/2$	---	---
	$f(x_i)$	0	$\frac{3}{2}$	1	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$
R	x_i	2π	---	---	---	---	---
	$f(x_i)$	2	---	---	---	2	0

Threshold	Sum S
$\pi/4$	2
$3\pi/4$	3
$5\pi/4$	3
$7\pi/4$	2

For $t = \pi/4$, the right targets sort to $1/2, 1, 3/2, 2$. Their median is $5/4$, and their MAE is $(3/4 + 1/4 + 1/4 + 3/4)/4 = 1/2$.

The thresholds $\pi/4$ and $7\pi/4$ tie. The prescribed rule selects $\pi/4$:

The single stump and the bootstrap trees are shown together in Solution 4.3.

$$\hat{f}_{\text{single}}(x) = \begin{cases} 0 & x < \pi/4 \\ 5/4 & x \geq \pi/4 \end{cases}$$

Its training MAE is $2/5$. The unsplit targets sort to $0, 1/2, 1, 3/2, 2$, with median 1 and MAE $(1 + 1/2 + 0 + 1/2 + 1)/5 = 3/5$.

Solution 4.3 - Train three trees

Sample A

Threshold $t = \pi/4$

Child		1	2	3	Median	MAE
L	x_i	0	0	---	---	---
	$f(x_i)$	0	0	---	0	0
R	x_i	$\pi/2$	π	2π	---	---
	$f(x_i)$	$\frac{3}{2}$	1	2	$\frac{3}{2}$	$\frac{1}{3}$

Threshold $t = 3\pi/4$

Child		1	2	3	Median	MAE
L	x_i	0	0	$\pi/2$	---	---
	$f(x_i)$	0	0	$\frac{3}{2}$	0	$\frac{1}{2}$
R	x_i	π	2π	---	---	---
	$f(x_i)$	1	2	---	$\frac{3}{2}$	$\frac{1}{2}$

Threshold $t = 3\pi/2$

Child		1	2	3	4	Median	MAE
L	x_i	0	0	$\pi/2$	π	---	---
	$f(x_i)$	0	0	$\frac{3}{2}$	1	$\frac{1}{2}$	$\frac{5}{8}$
R	x_i	2π	---	---	---	---	---
	$f(x_i)$	2	---	---	---	2	0

Threshold	Sum S
$\pi/4$	1
$3\pi/4$	$\frac{5}{2}$
$3\pi/2$	$\frac{5}{2}$

The best split gives

$$\hat{f}_A(x) = \begin{cases} 0 & x < \pi/4 \\ \frac{3}{2} & x \geq \pi/4 \end{cases}$$

Sample B

Threshold $t = \pi/2$

Child		1	2	3	4	Median	MAE
L	x_i	0	---	---	---	---	---
	$f(x_i)$	0	---	---	---	0	0
R	x_i	π	$3\pi/2$	2π	2π	---	---
	$f(x_i)$	1	$\frac{1}{2}$	2	2	$\frac{3}{2}$	$\frac{5}{8}$

Threshold $t = 5\pi/4$

Child		1	2	3	Median	MAE
L	x_i	0	π	---	---	---
	$f(x_i)$	0	1	---	$\frac{1}{2}$	$\frac{1}{2}$
R	x_i	$3\pi/2$	2π	2π	---	---
	$f(x_i)$	$\frac{1}{2}$	2	2	2	$\frac{1}{2}$

Threshold $t = 7\pi/4$

Child		1	2	3	Median	MAE
L	x_i	0	π	$3\pi/2$	---	---
	$f(x_i)$	0	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$
R	x_i	2π	2π	---	---	---
	$f(x_i)$	2	2	---	2	0

Threshold	Sum S
$\pi/2$	$\frac{5}{2}$
$5\pi/4$	$\frac{5}{2}$
$7\pi/4$	1

The best split gives

$$\hat{f}_B(x) = \begin{cases} \frac{1}{2} & x < 7\pi/4 \\ 2 & x \geq 7\pi/4 \end{cases}$$

Sample C

Threshold $t = \pi/2$

Child		1	2	3	Median	MAE
L	x_i	0	0	---	---	---
	$f(x_i)$	0	0	---	0	0
R	x_i	π	π	2π	---	---
	$f(x_i)$	1	1	2	1	$\frac{1}{3}$

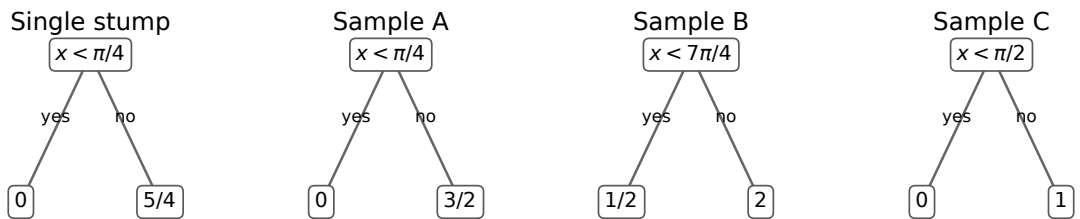
Threshold $t = 3\pi/2$

Child		1	2	3	4	Median	MAE
L	x_i	0	0	π	π	---	---
	$f(x_i)$	0	0	1	1	$\frac{1}{2}$	$\frac{1}{2}$
R	x_i	2π	---	---	---	---	---
	$f(x_i)$	2	---	---	---	2	0

Threshold	Sum S
$\pi/2$	1
$3\pi/2$	2

The best split gives

$$\hat{f}_C(x) = \begin{cases} 0 & x < \pi/2 \\ 1 & x \geq \pi/2 \end{cases}$$



Left branches satisfy the condition; right branches do not.

Solution 4.4 - Combine the predictions

x	Target $f(x)$	Tree A	Tree B	Tree C	Forest	Absolute forest error
0	0	0	$\frac{1}{2}$	0	$\frac{1}{6}$	$\frac{1}{6}$
$\pi/2$	$\frac{3}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$
π	1	$\frac{3}{2}$	$\frac{1}{2}$	1	1	0
$3\pi/2$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$
2π	2	$\frac{3}{2}$	2	1	$\frac{3}{2}$	$\frac{1}{2}$

Over $[0, 2\pi]$,

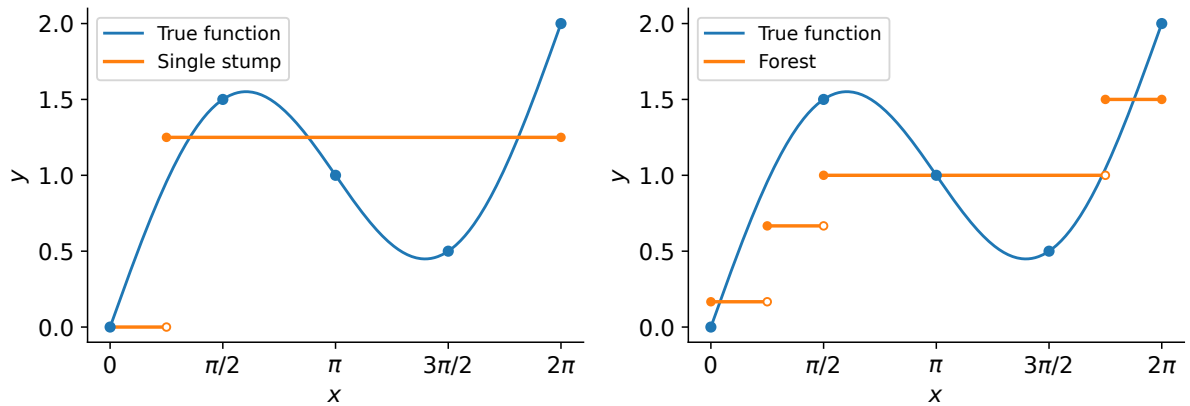
$$\hat{f}_{\text{forest}}(x) = \begin{cases} 1/6 & 0 \leq x < \pi/4 \\ 2/3 & \pi/4 \leq x < \pi/2 \\ 1 & \pi/2 \leq x < 7\pi/4 \\ 3/2 & 7\pi/4 \leq x \leq 2\pi \end{cases}$$

The plot below uses the value of the segment to the right at each threshold.

The MAE on the original five observations is

$$\text{MAE}_{\text{forest}} = \frac{1/6 + 1/2 + 0 + 1/2 + 1/2}{5} = \frac{1}{3}$$

This is smaller than the comparison tree's MAE of $2/5$, by $1/15$.



Solutions - Conceptual questions

1. With squared error, a leaf predicts the **mean target value** of the samples in that leaf.
2. The mean minimizes the sum of squared residuals.
3. A split is chosen to minimize the total error in its child nodes, or equivalently to maximize the reduction in error.
4. Candidate thresholds can be placed between neighboring distinct sorted feature values.
5. Every leaf predicts a constant, so the overall prediction is piecewise constant.
6. Increasing depth creates more and smaller regions and usually reduces training error.
7. A deep tree can adapt to random fluctuations in the training sample.
8. With several input features, the tree must choose both the **feature** and the **threshold**.
9. The regions are rectangles, possibly unbounded.
10. Bootstrap samples repeat some observations and omit others. This changes child medians, split errors and sometimes the candidate thresholds. Every repetition counts separately when sorting targets and calculating errors.
11. No. A finite average of piecewise-constant tree predictions is still piecewise constant. Its possible changes occur at the constituent trees' thresholds. Averaging can reduce jumps, but does not guarantee continuity; in this example the forest has three nonzero jumps.
12. No. The training error measures accuracy only at the observed positions. A smaller training error does not guarantee more accurate predictions between them.

Key concepts

Concept	Meaning in this exercise
Regression	Predicting a numerical target value.
Observation / feature / target	One data point / an input value / the value to be predicted.
Training data	Observations used to choose splits and leaf predictions.
Regression tree	A sequence of splits dividing the input space into regions with constant predictions.
Node / root node	A stage containing observations / the initial node containing all training observations.
Parent / child node	A node before a split / one of the nodes produced by that split.
Internal node / leaf	A node that splits / a final node assigning a constant prediction.
Split / threshold	A division such as $x < t$ versus $x \geq t$ / its boundary t .
Candidate split	A feature-threshold choice tested during training; both children must be non-empty.
Residual	Signed prediction error $r_i = y_i - \hat{y}_i$.
Loss	A numerical measure of prediction error.
SSE	Sum of squared residuals.
MSE / RMSE	Mean squared residual / its square root, in target units.
MAE	Mean absolute residual.
Mean / median	Arithmetic average / middle of the sorted values, conventionally averaging the middle pair for even sample size.
Error reduction	Parent SSE minus the sum of child SSEs.
Tree depth	Maximum number of splits on a root-to-leaf path.
Piecewise-constant prediction	A prediction constant inside each leaf region.
Training error	Prediction error on the observations used to build the model.
Noise / overfitting	Random target variation / fitting sample-specific details in a way that harms prediction on new data.
Feature space	The space of possible combinations of input features.
Axis-aligned split	A split testing only one feature.
Bootstrap sample	A sample drawn with replacement; observations may repeat or be absent. Here each sample has five draws.
Forest	A collection of trees whose predictions are combined; here by their arithmetic mean.
Total absolute split error	$S = N_L \text{MAE}_L + N_R \text{MAE}_R$, counting each observation in its child.