Quiz week 3: tree-based methods

Q1

Which of the following statements are correct about classification and regression trees (CART)?

1. One advantage of CART is smoothness: small perturbations in the input data do not dramatically change the response

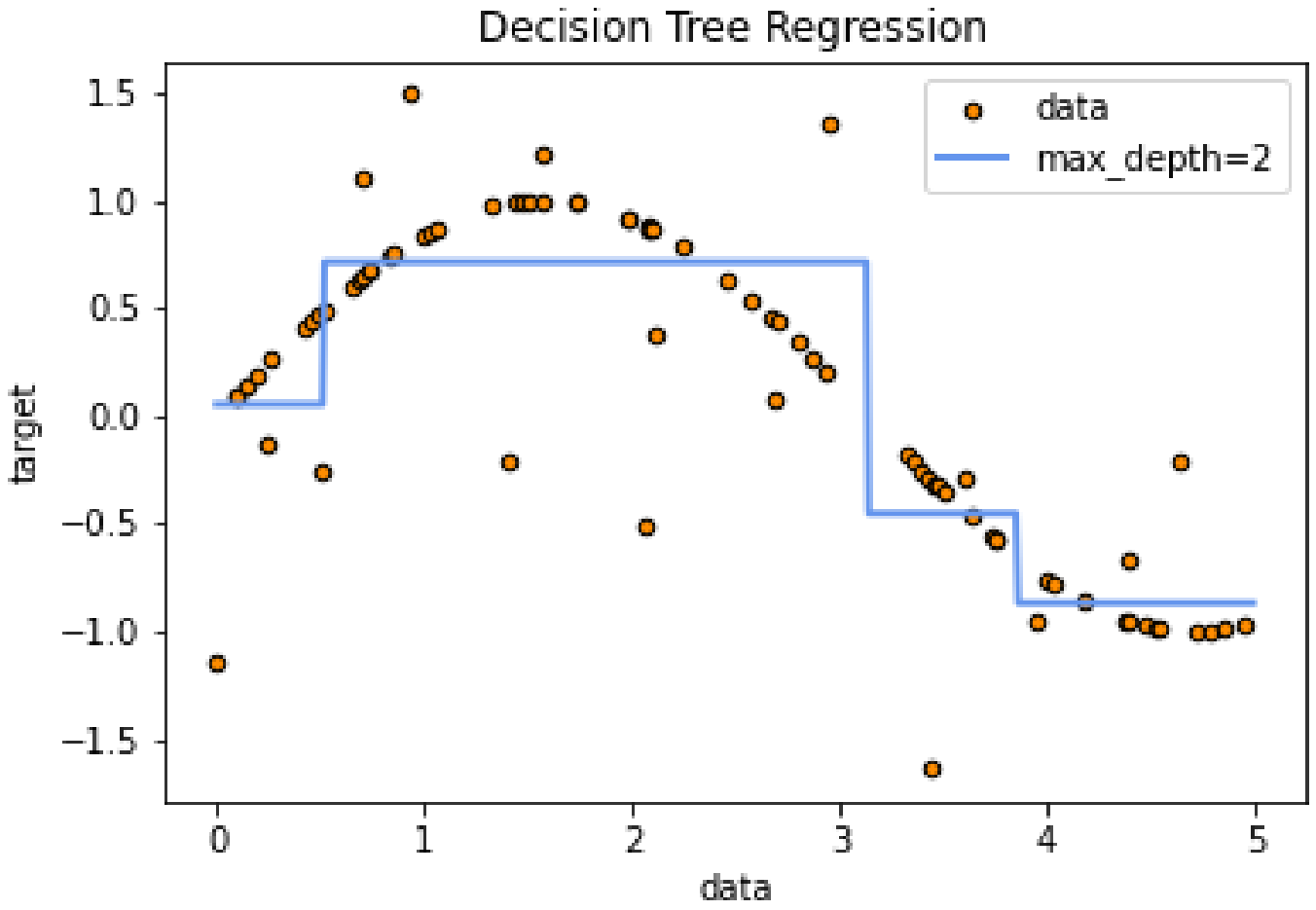
2. One advantage of CART is interpretability: it is easy to understand which features learnt generated the predictions

3. One advantage of CART is flexibility: no assumptions of data distribution and no transformations needed

4. One disadvantage of CART is overfitting: they do not easily generalise to new unseen data

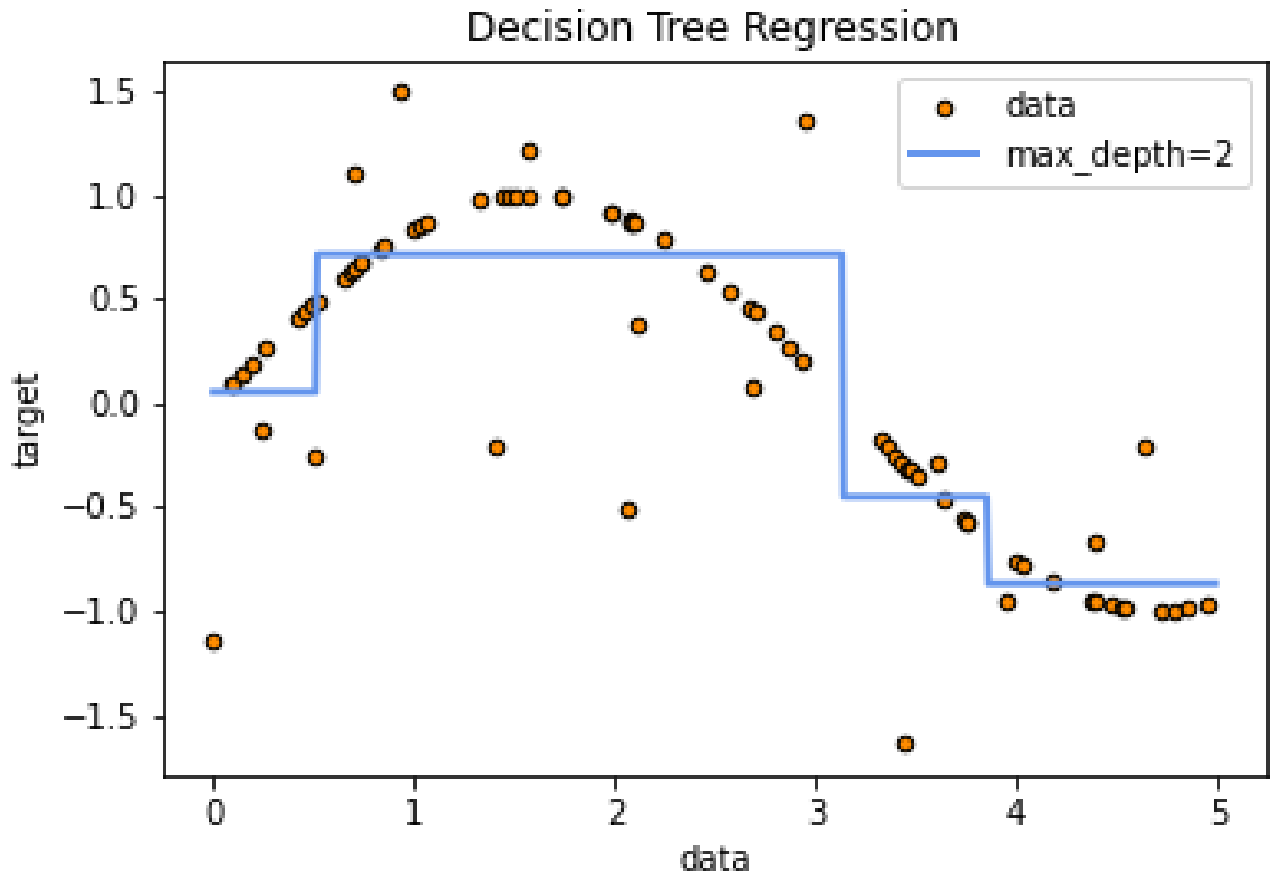
Q1

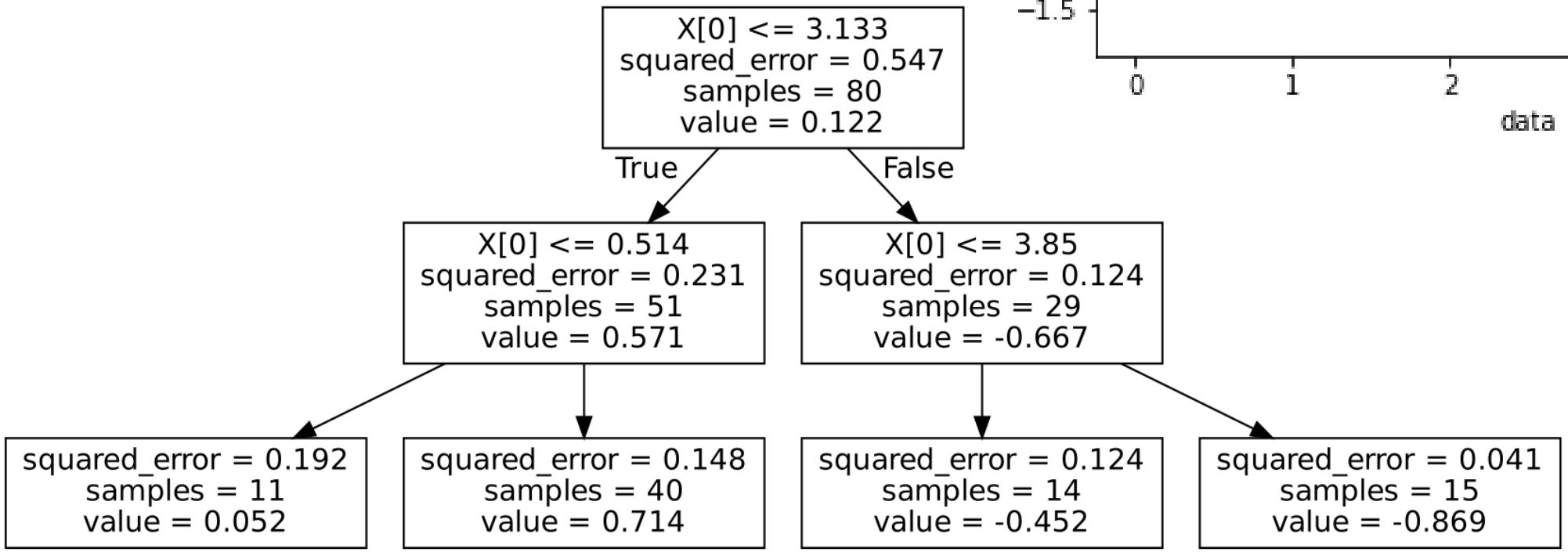
3. One *advantage* of CART is *flexibility*: no assumptions of data distribution and no transformations needed -> True, here are fitting a non-linear function using CART



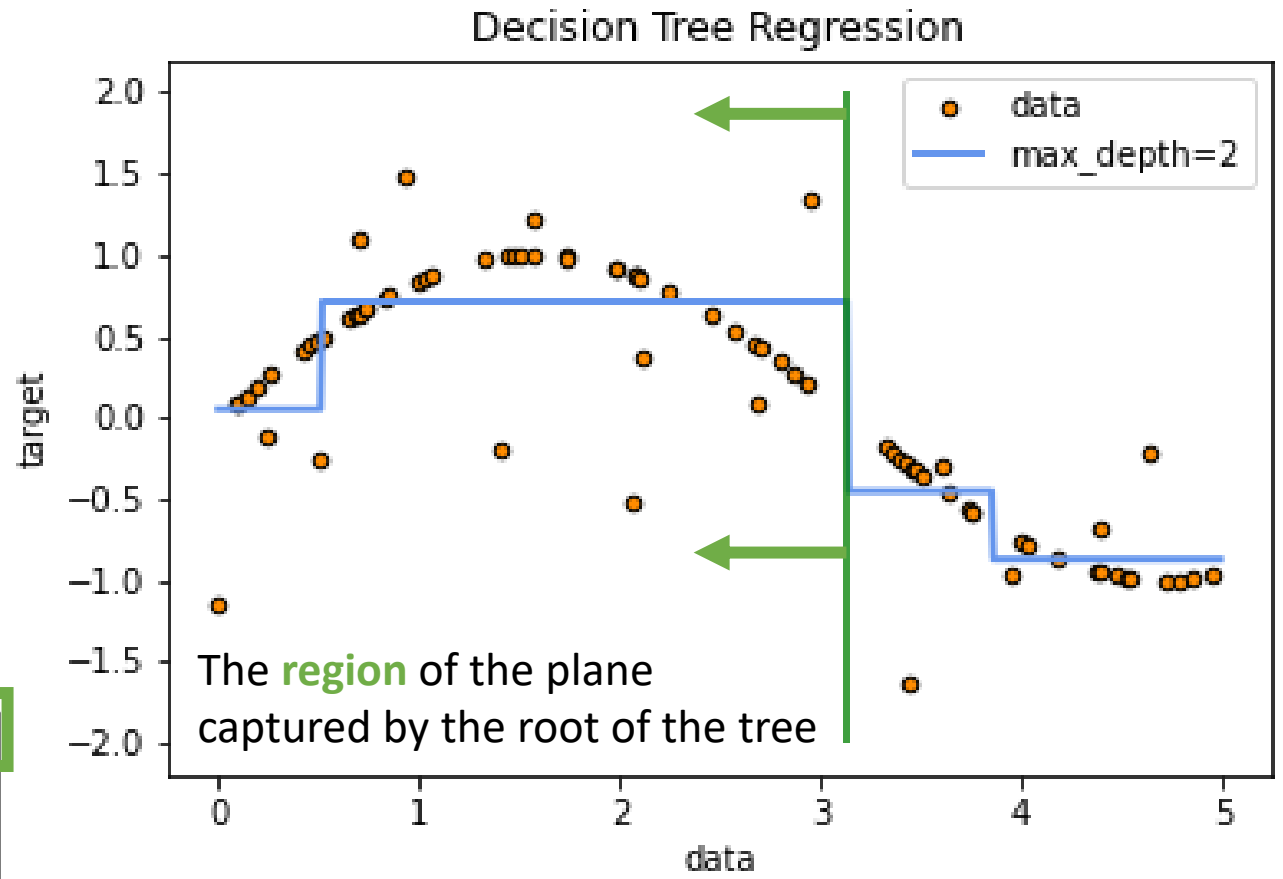
Q1

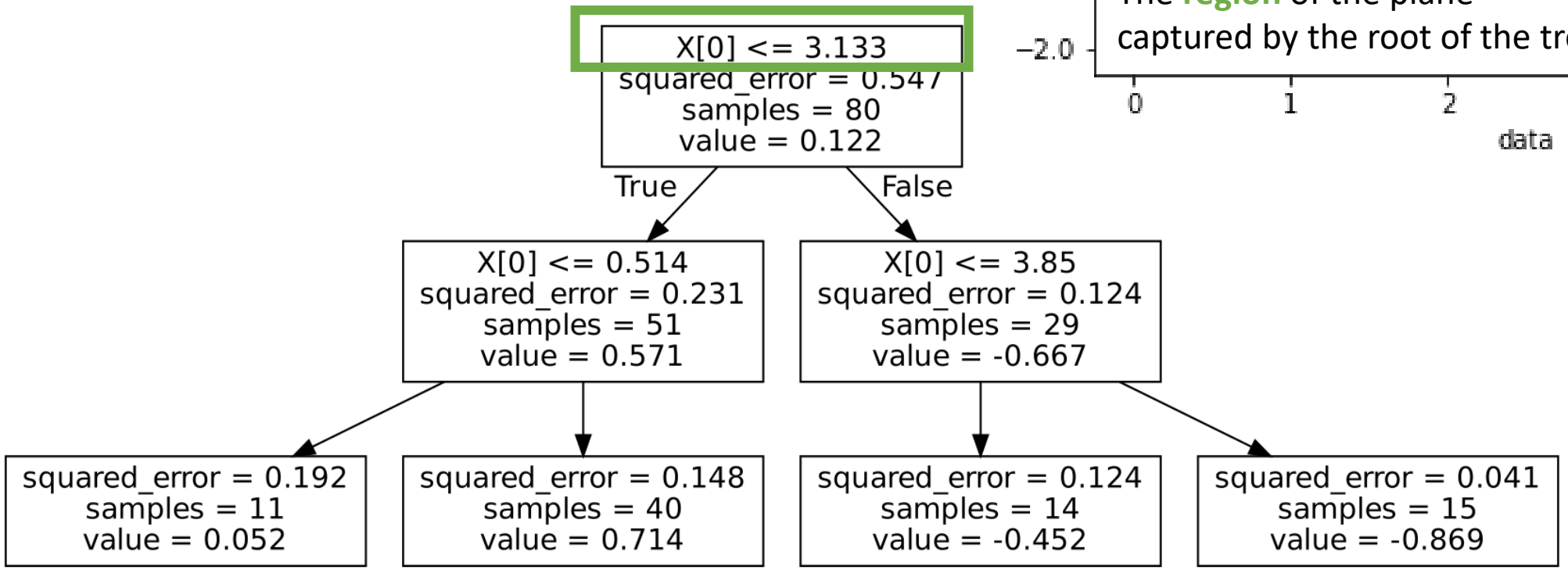
How can we visualise a regression tree?





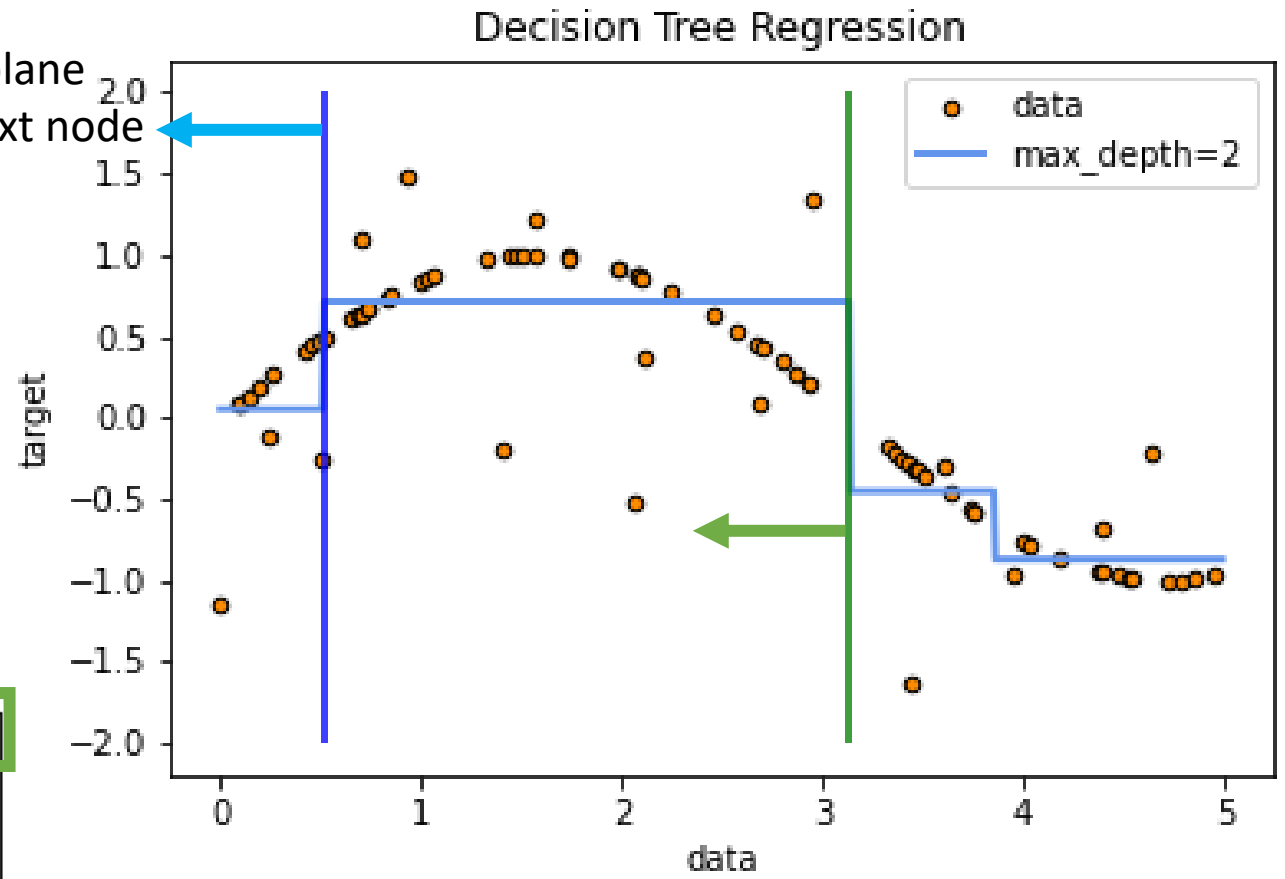
Q1

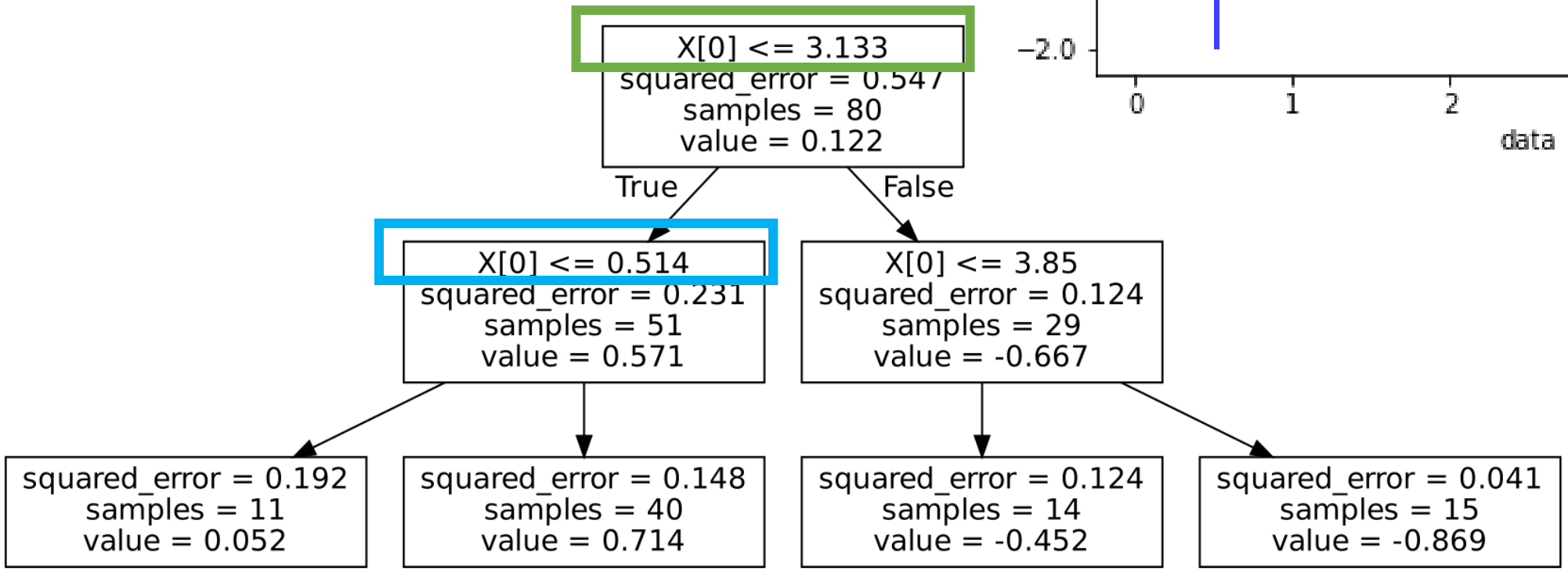

Decision Tree Regression
data
max_depth=2
target
data
The region of the plane
captured by the root of the tree


X[0] <= 3.133
squared_error = 0.547
samples = 80
value = 0.122
True
False
X[0] <= 0.514
squared_error = 0.231
samples = 51
value = 0.571
X[0] <= 3.85
squared_error = 0.124
samples = 29
value = -0.667
squared_error = 0.192
samples = 11
value = 0.052
squared_error = 0.148
samples = 40
value = 0.714
squared_error = 0.124
samples = 14
value = -0.452
squared_error = 0.041
samples = 15
value = -0.869

Q1

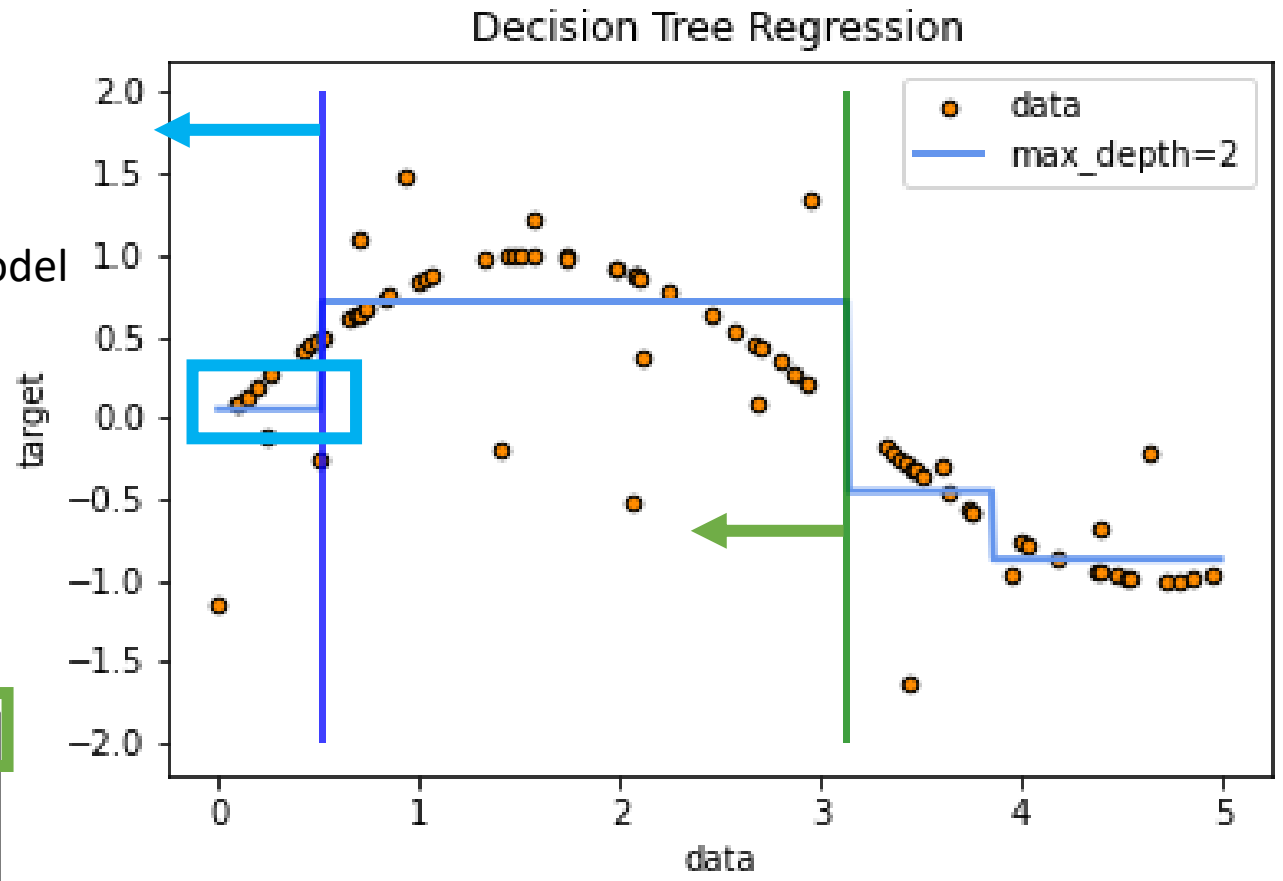
The region of the plane captured by the next node

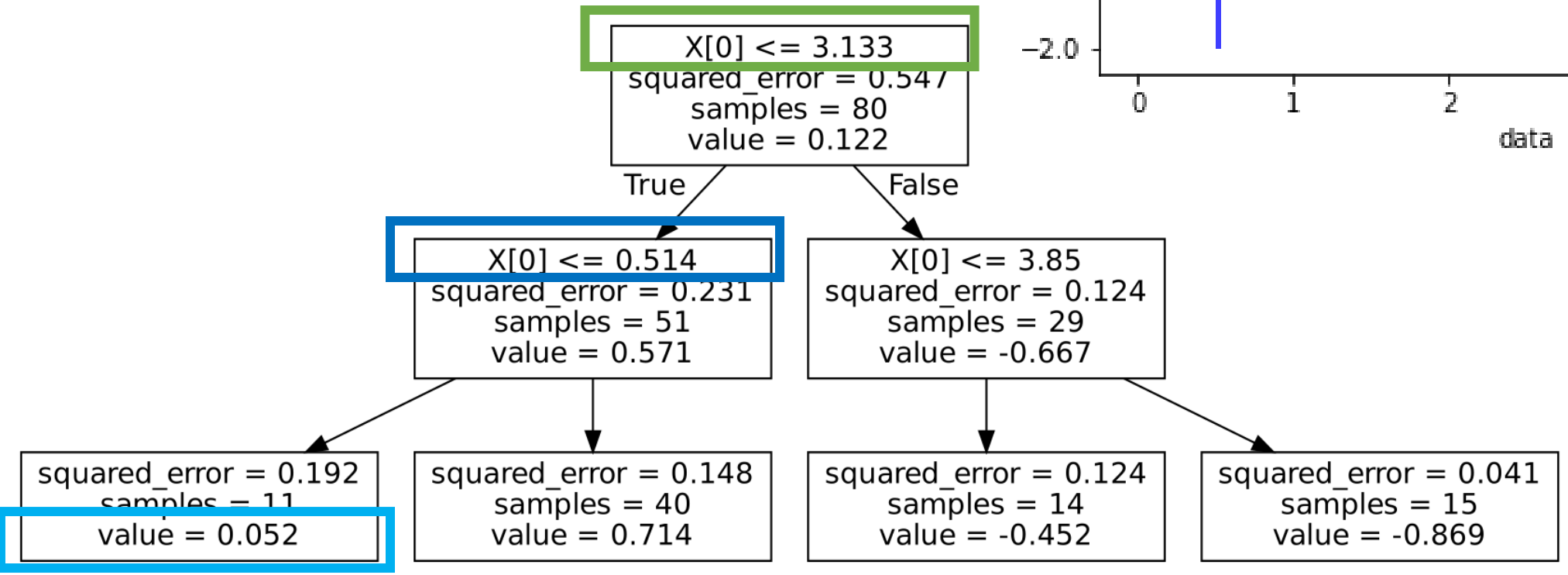




Q1

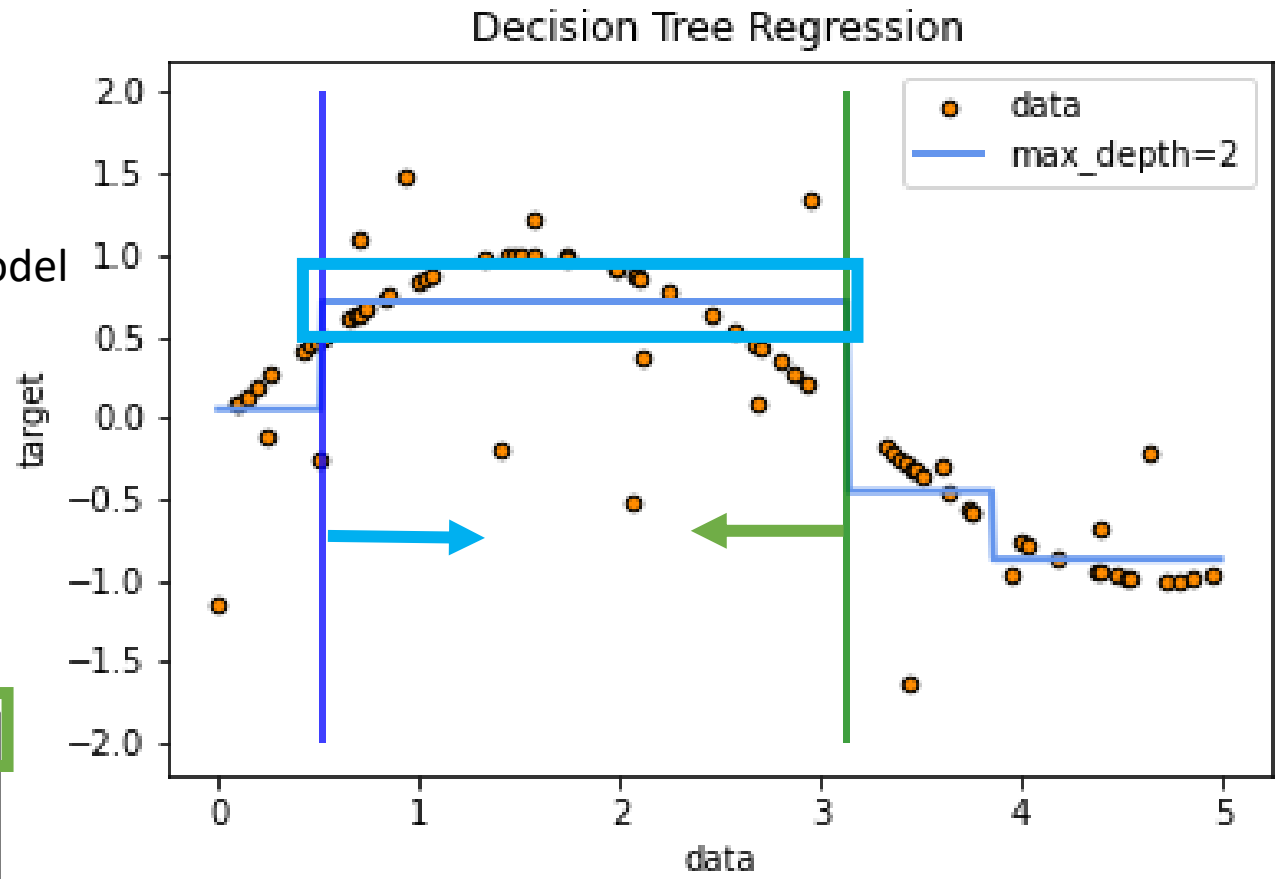
The **value** inferred by the model for this region of the plane

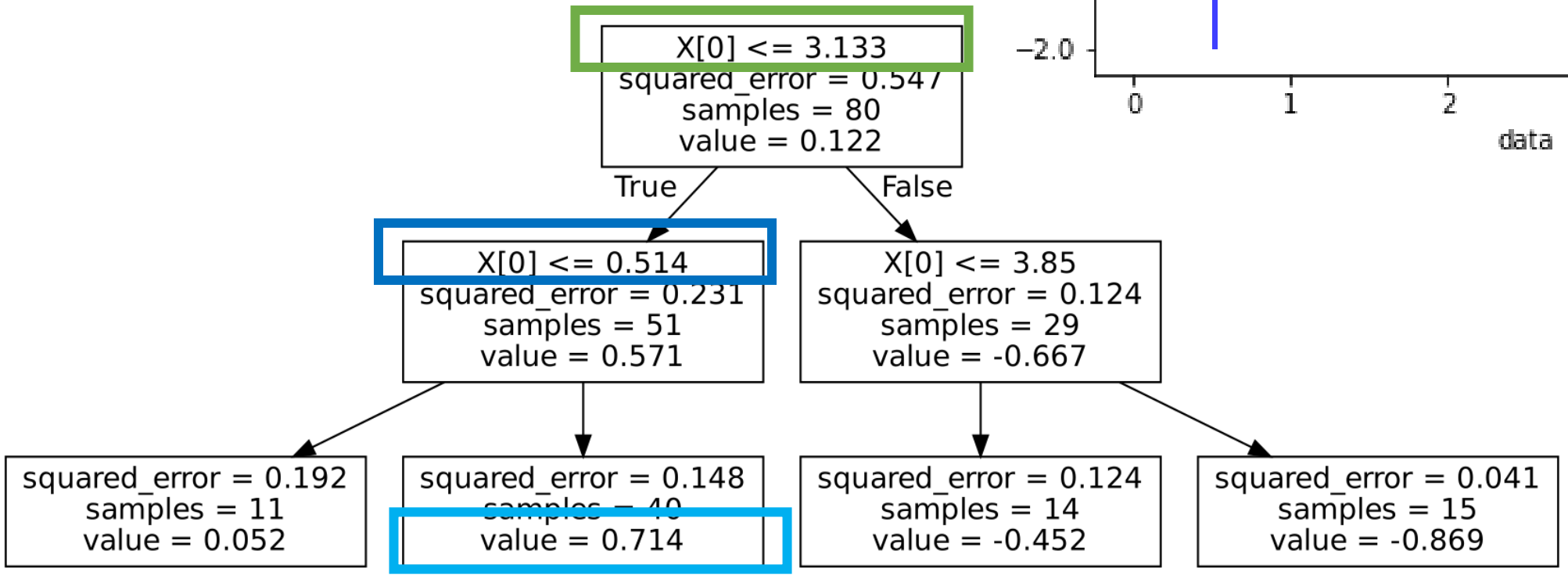




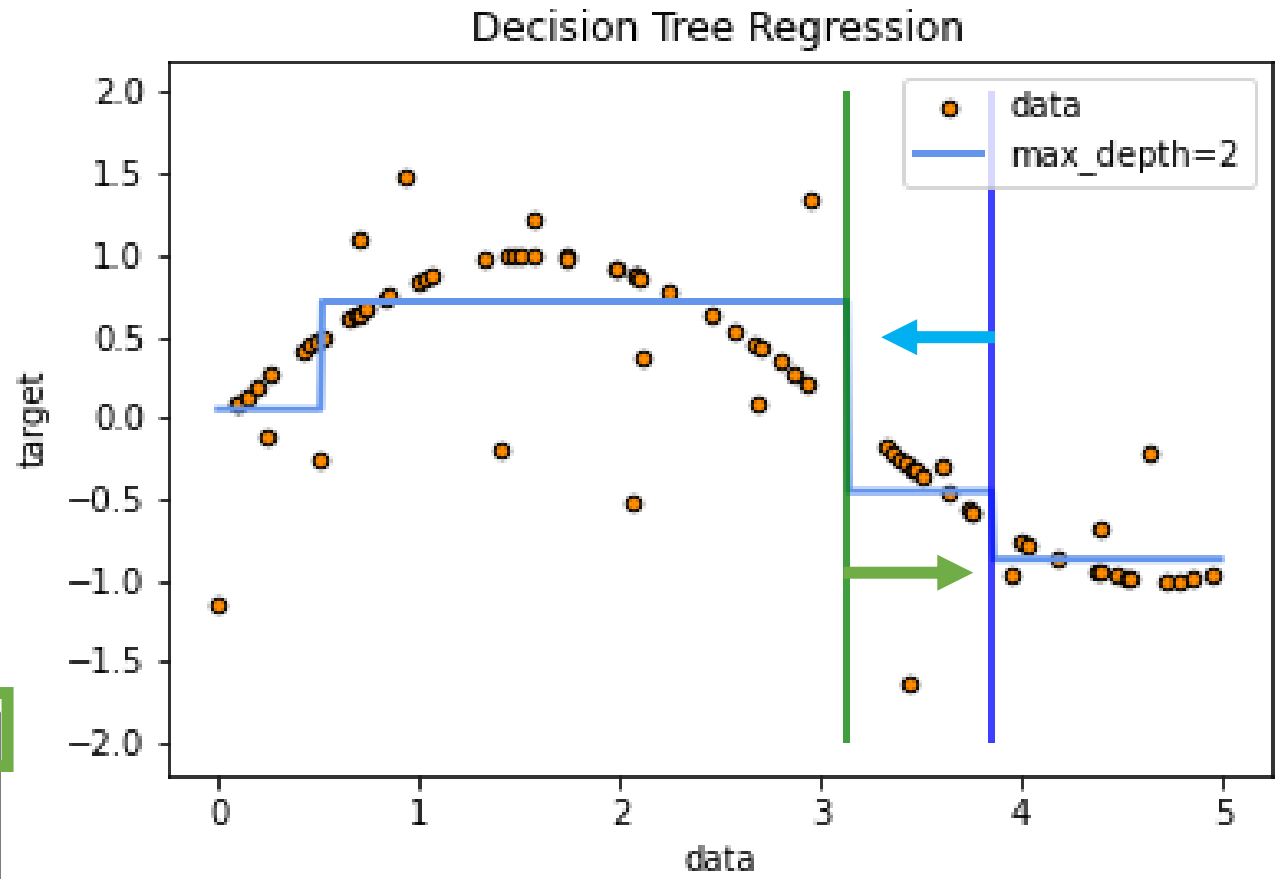
Q1

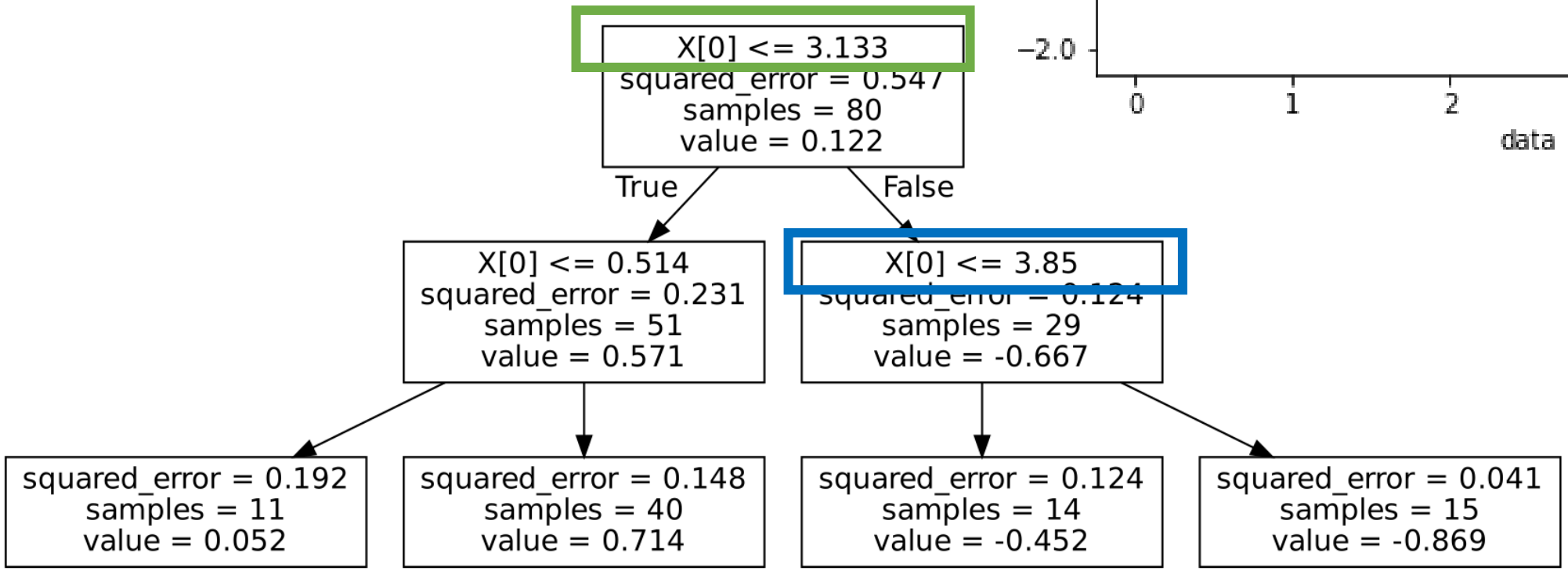
The **value** inferred by the model for this region of the plane





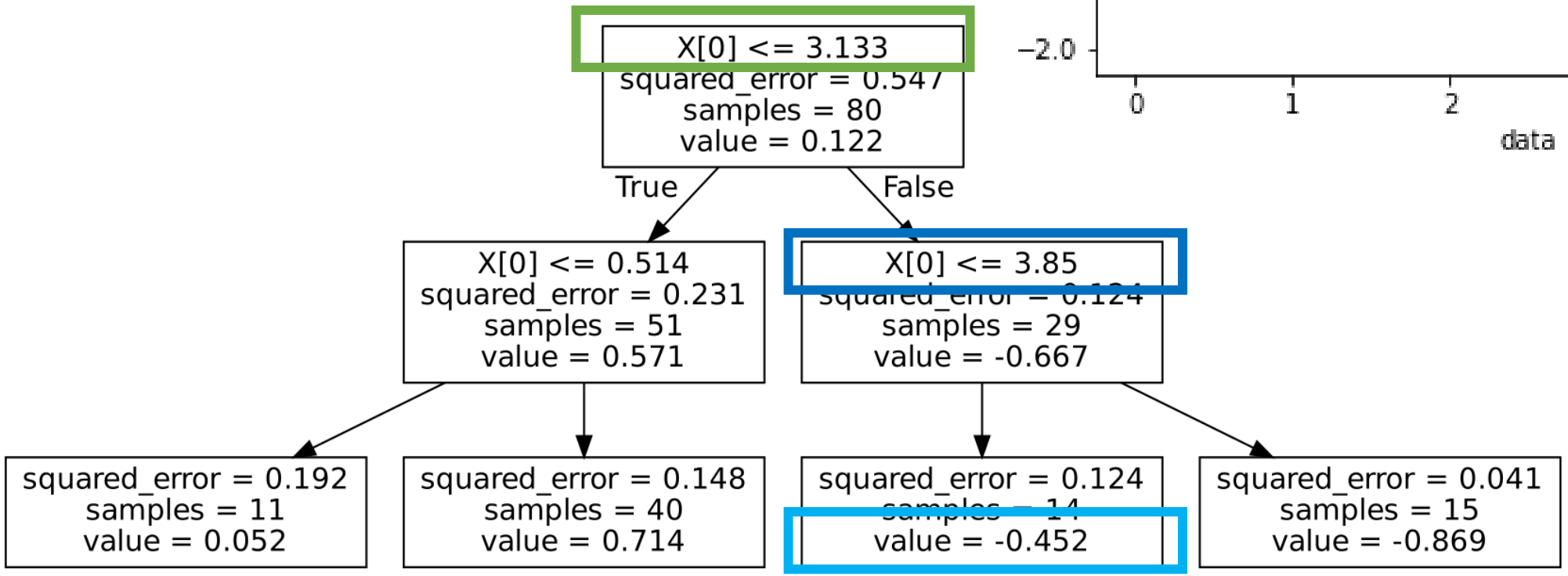
Q1


Decision Tree Regression
data
max_depth=2
target
data
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
1
2
3
4
5


X[0] <= 3.133
squared_error = 0.547
samples = 80
value = 0.122
True
False
X[0] <= 0.514
squared_error = 0.231
samples = 51
value = 0.571
X[0] <= 3.85
samples = 29
value = -0.667
squared_error = 0.192
samples = 11
value = 0.052
squared_error = 0.148
samples = 40
value = 0.714
squared_error = 0.124
samples = 14
value = -0.452
squared_error = 0.041
samples = 15
value = -0.869

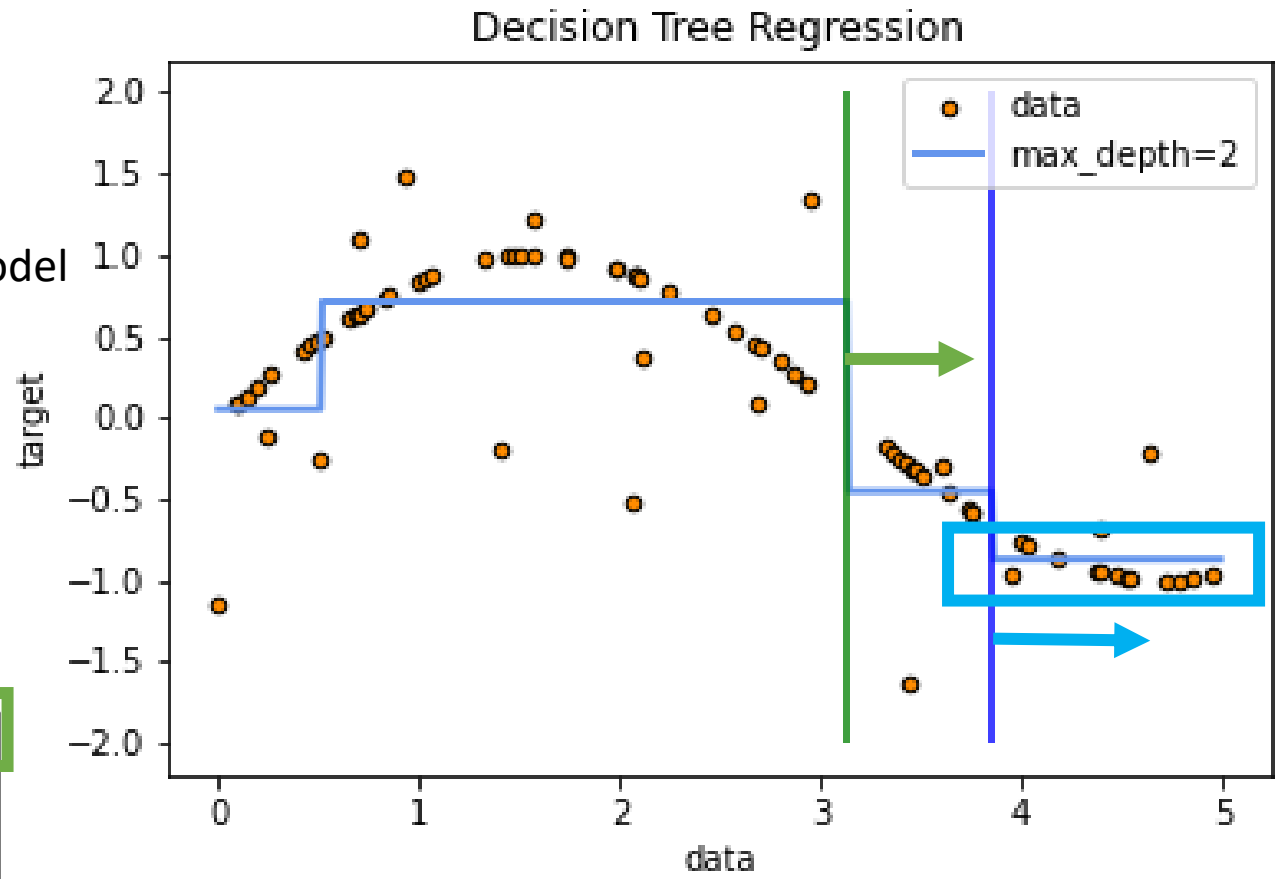
Q1

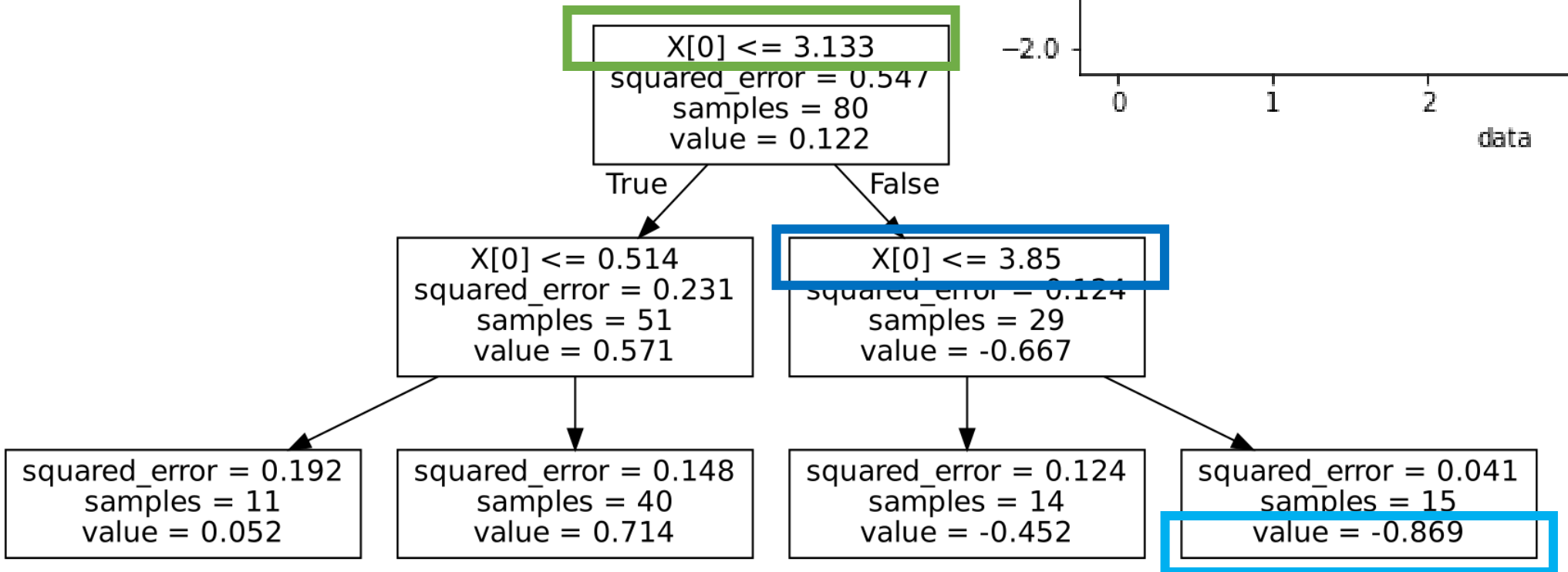
The **value** inferred by the model for this region of the plane



Q1

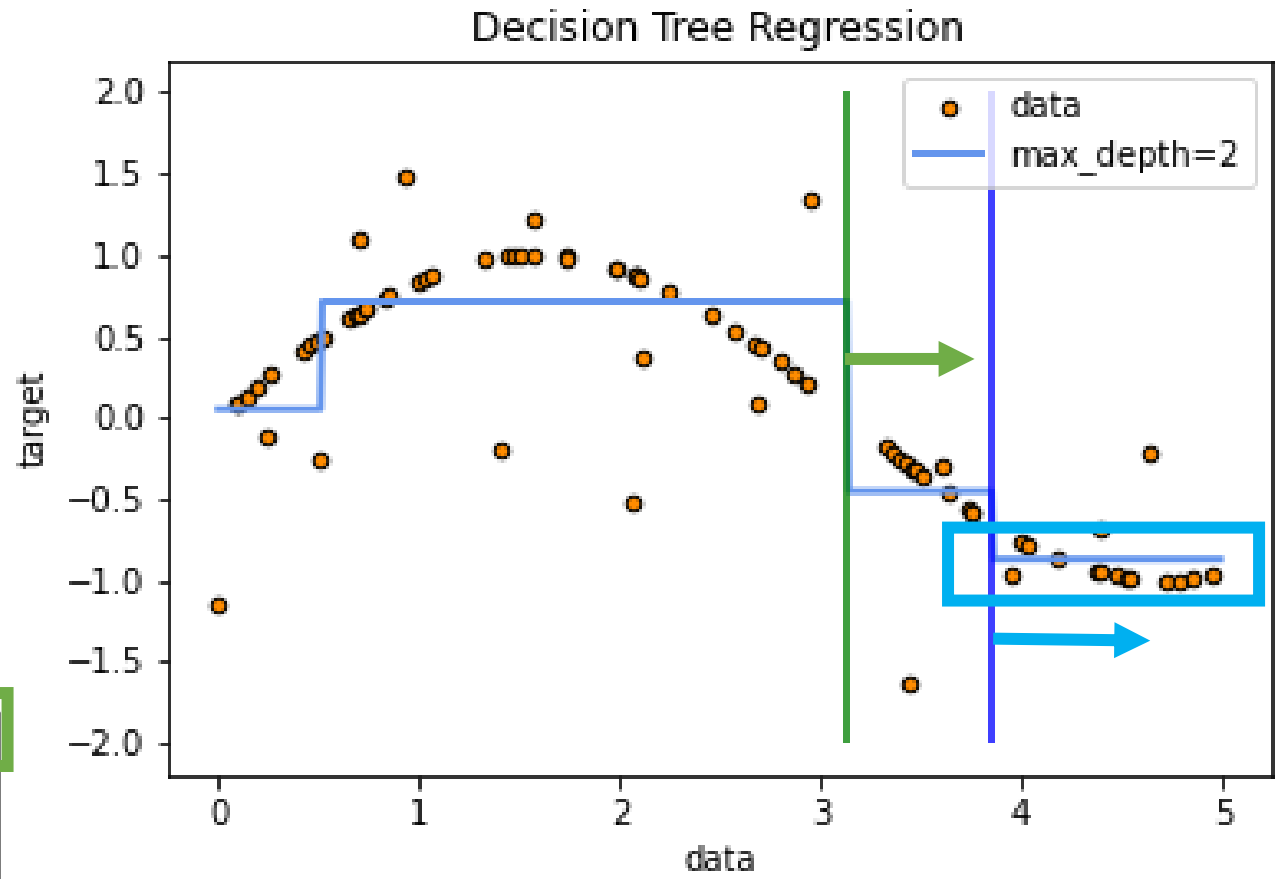
The **value** inferred by the model for this region of the plane

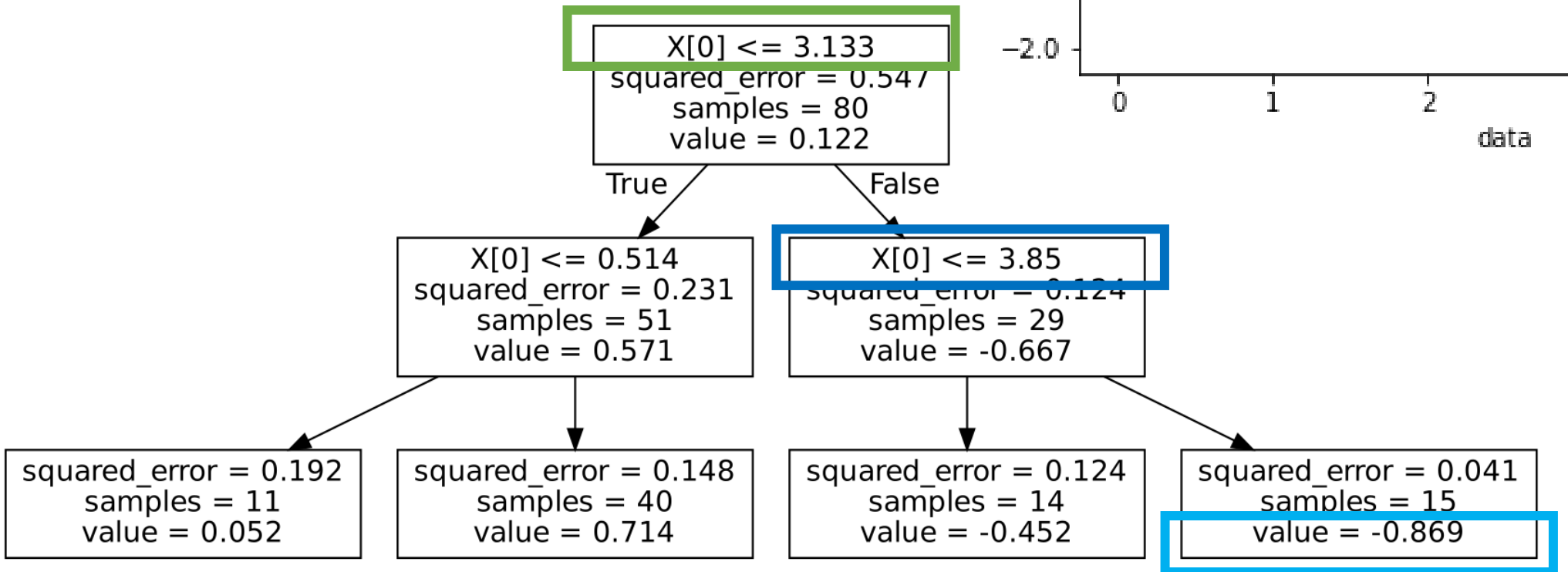




Q1

2. One *advantage* of CART is *interpretability*: it is easy to understand which features learnt generated the predictions -> True

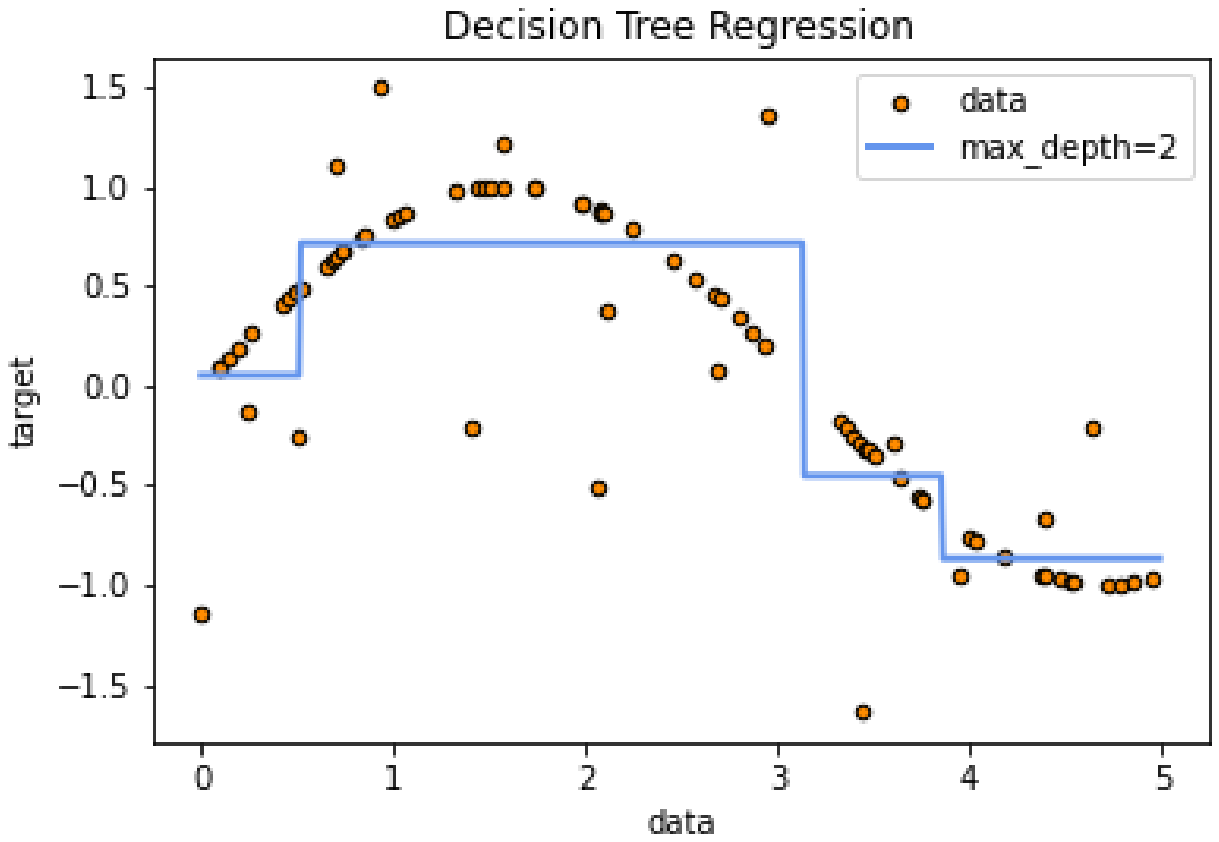




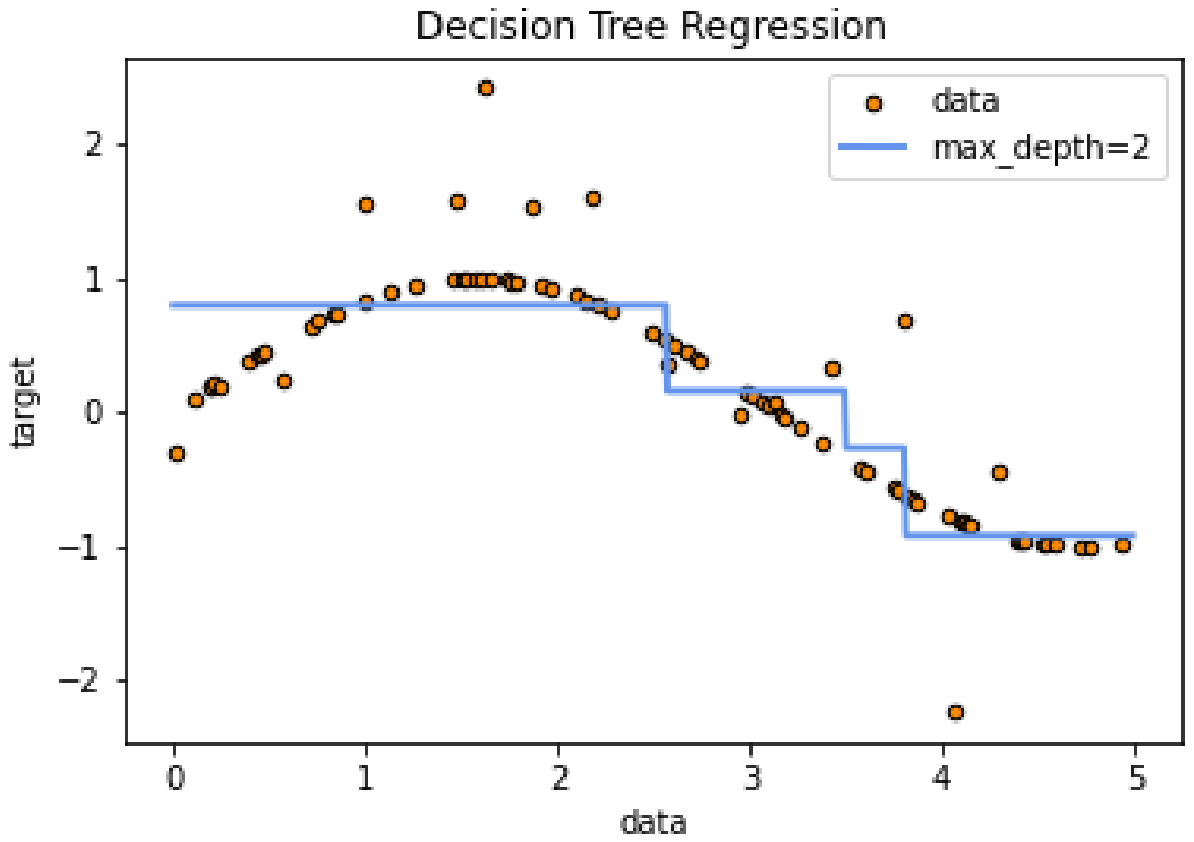
Q1

1. One *advantage* of CART is *smoothness*: small perturbations in the input data do not dramatically change the response -> False, here we have generated the same noisy function with a different seed and we got a completely different regression estimation

Seed = 1



Seed = 10



Q2

Classification and regression trees (CART) have hyper-parameters. Which of the following statements are correct?

1. **CART's hyper-parameters include the maximal depth of the tree and the minimal number of records on a node**
2. **CART's hyper-parameters represent a trade-off between performance and overfitting and are user-defined, though can be tuned by cross-validation**
3. The values of the hyper-parameters are inferred from the data via the learning process (training)

Adapted from this source: https://www.vebuso.com/2020/01/decision-tree-intuition-from-concept-to-application

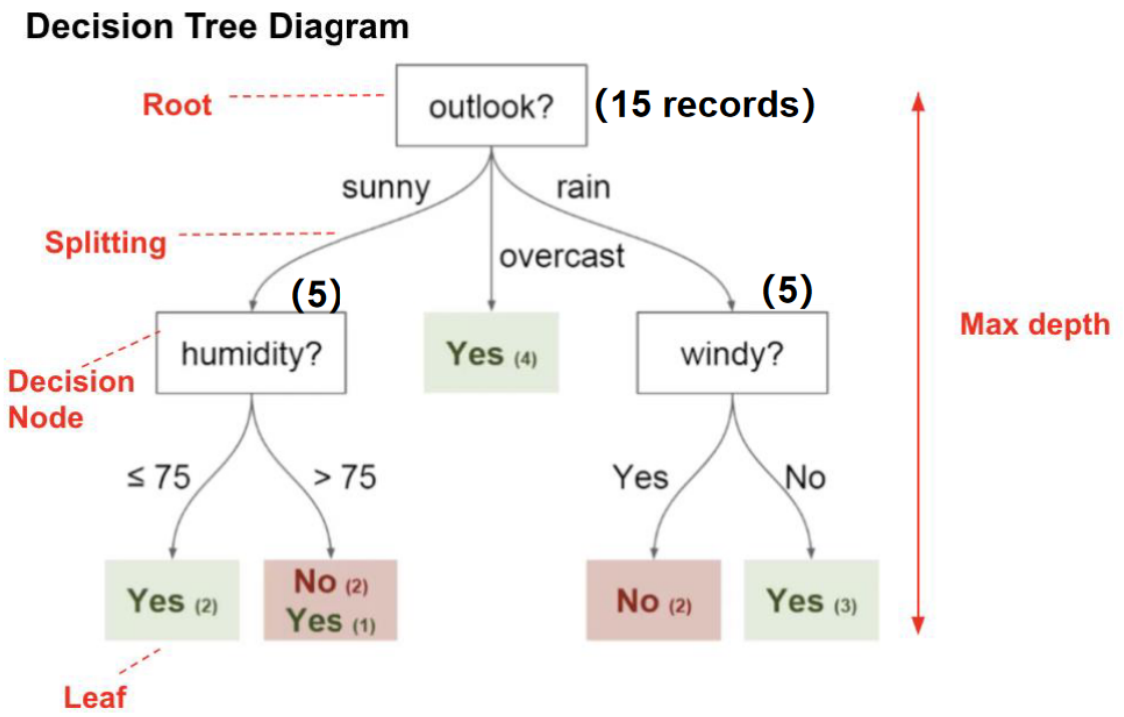
Q2

Classification and regression trees (CART) have hyper-parameters. Which of the following statements are correct?

1. CART's hyper-parameters include the maximal depth of the tree and the minimal number of records on a node -> True

Dataset

Temperature	Outlook	Humidity	Windy	Played?
Mild	Sunny	80	No	Yes
Hot	Sunny	75	Yes	No
Hot	Overcast	77	No	Yes
Cool	Rain	70	No	Yes
Cool	Overcast	72	Yes	Yes
Mild	Sunny	77	No	No
Cool	Sunny	70	No	Yes
Mild	Rain	69	No	Yes
Mild	Sunny	65	Yes	Yes
Mild	Overcast	77	Yes	Yes
Hot	Overcast	74	No	Yes
Mild	Rain	77	Yes	No
Cool	Rain	73	Yes	No
Mild	Rain	78	No	Yes



Adapted from this source: https://www.vebuso.com/2020/01/decision-tree-intuition-from-concept-to-application

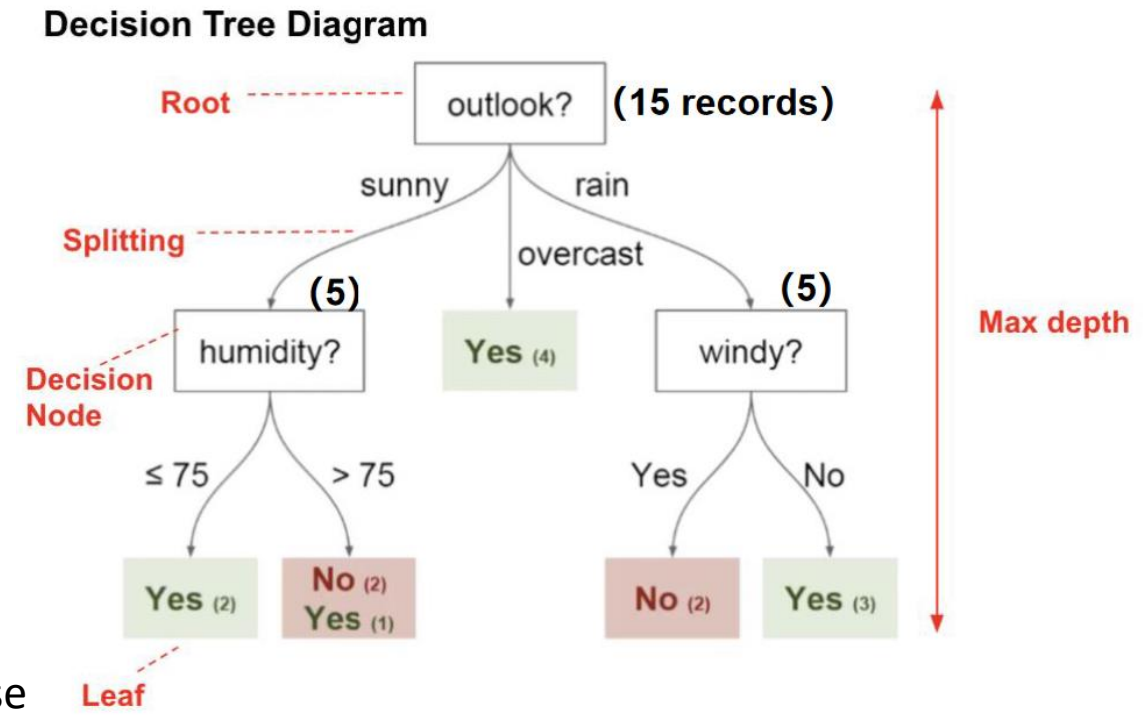
Q2

Classification and regression trees (CART) have hyper-parameters. Which of the following statements are correct?

1. CART's hyper-parameters include the maximal depth of the tree and the minimal number of records on a node -> True

The hyperparameters of a decision tree are:

max_depth, *min_samples_per_leaf*, *min_samples_split*, *max_leaf_nodes*, and *min_impurity_decrease*



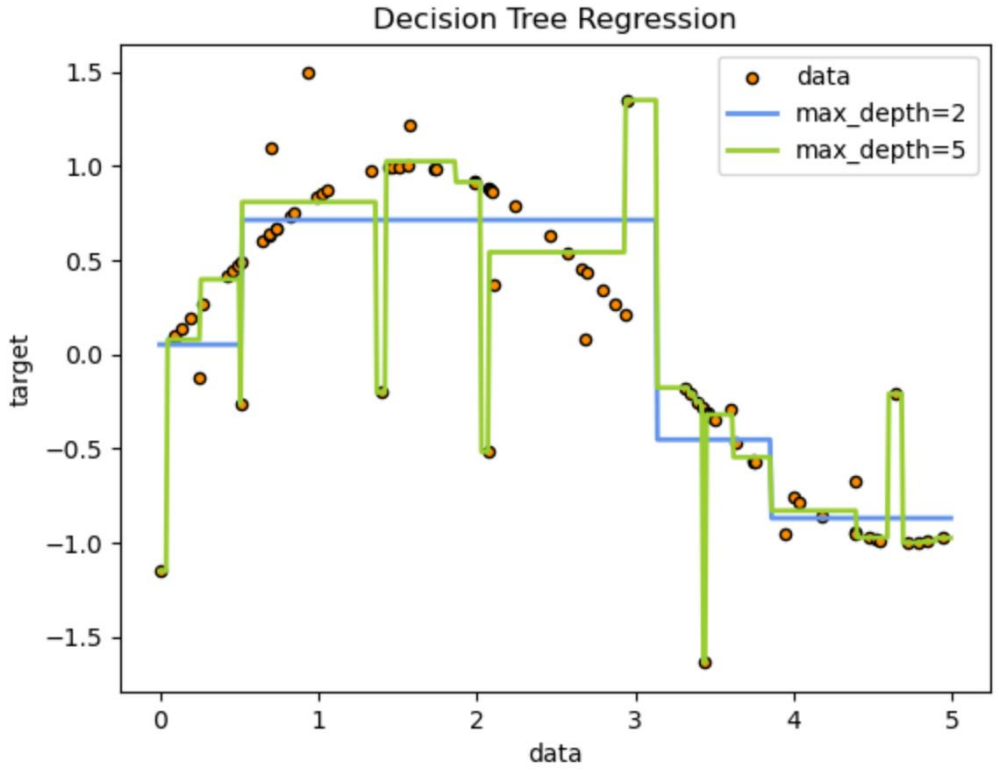
https://scikit-learn.org/stable/modules/tree.html#tips-on-practical-use

Adapted from this source: https://www.vebuso.com/2020/01/decision-tree-intuition-from-concept-to-application

Q2

Classification and regression trees (CART) have hyper-parameters. Which of the following statements are correct?

2. CART's hyper-parameters represent a trade-off between performance and overfitting and are user-defined, though can be tuned by cross-validation -> True

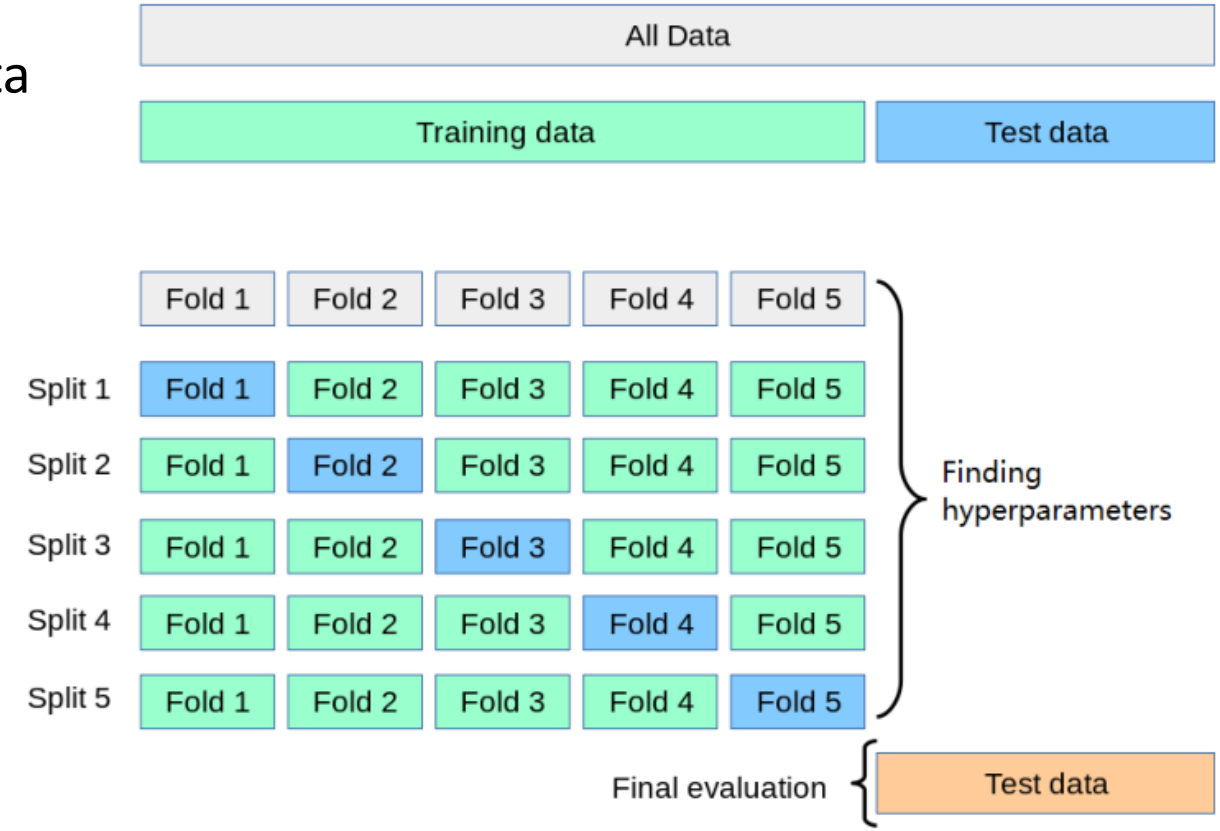


With greater max_depth, we can model more regions of the plane and increase the model's complexity

Q2

Classification and regression trees (CART) have hyper-parameters. Which of the following statements are correct?

3. The values of the hyper-parameters are inferred from the data via the learning process (training) -> False, they are inferred via cross-validation



Q3

3. CART are usually used as the base predictors of random forest (RF). Which of the following are correct?

1. RF constructs an ensemble of trees in parallel

2. RF repeatedly samples datapoints and features during training producing different trees

3. RF can be used for both classification and regression tasks

Q3

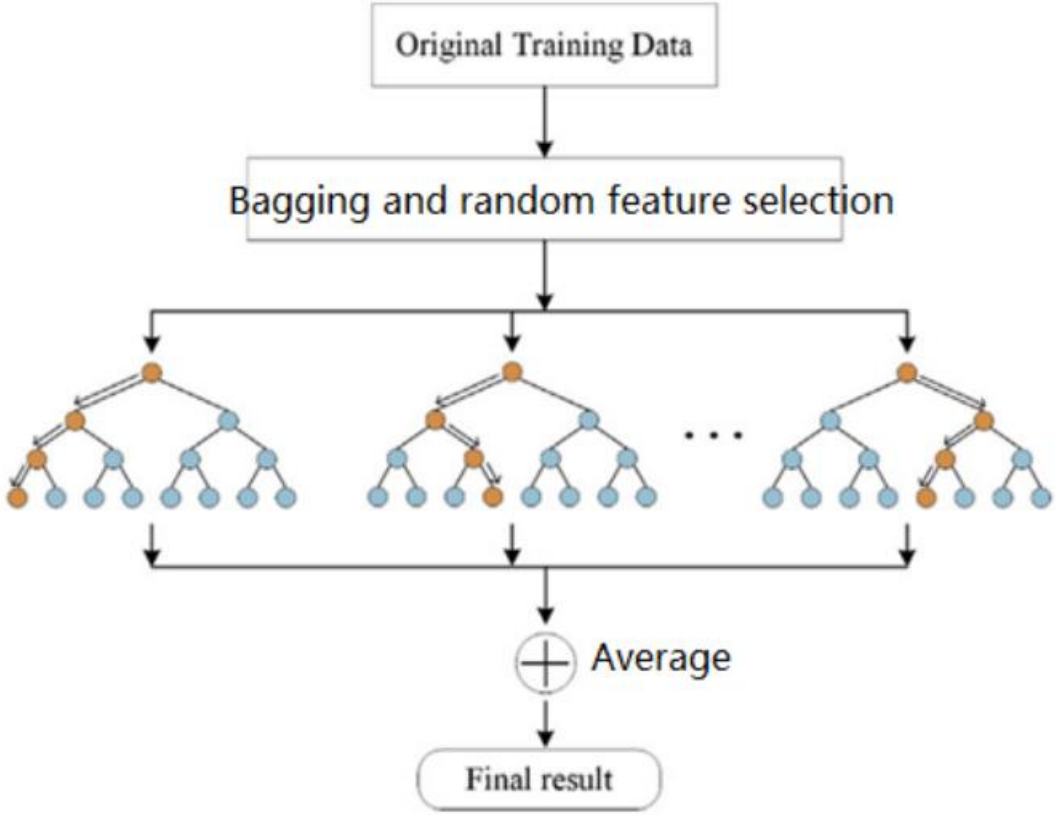
3. CART are usually used as the base predictors of random forest (RF). Which of the following are correct?

1. RF constructs an ensemble of trees in parallel -> True

2. RF repeatedly samples datapoints and features during training producing different trees -> True

3. RF can be used for both classification and

regression tasks -> True



Q4

4. CART are usually used as the base predictors of gradient-boosted decision trees (GBDT). Which of the following are correct?

1. GBDT constructs an ensemble of trees in parallel

2. During GBDT's fitting, a new CART predictor is trained using the residual from the last CART as the weight, considering the largest residuals

3. GBDT has been used successfully in many data science competitions

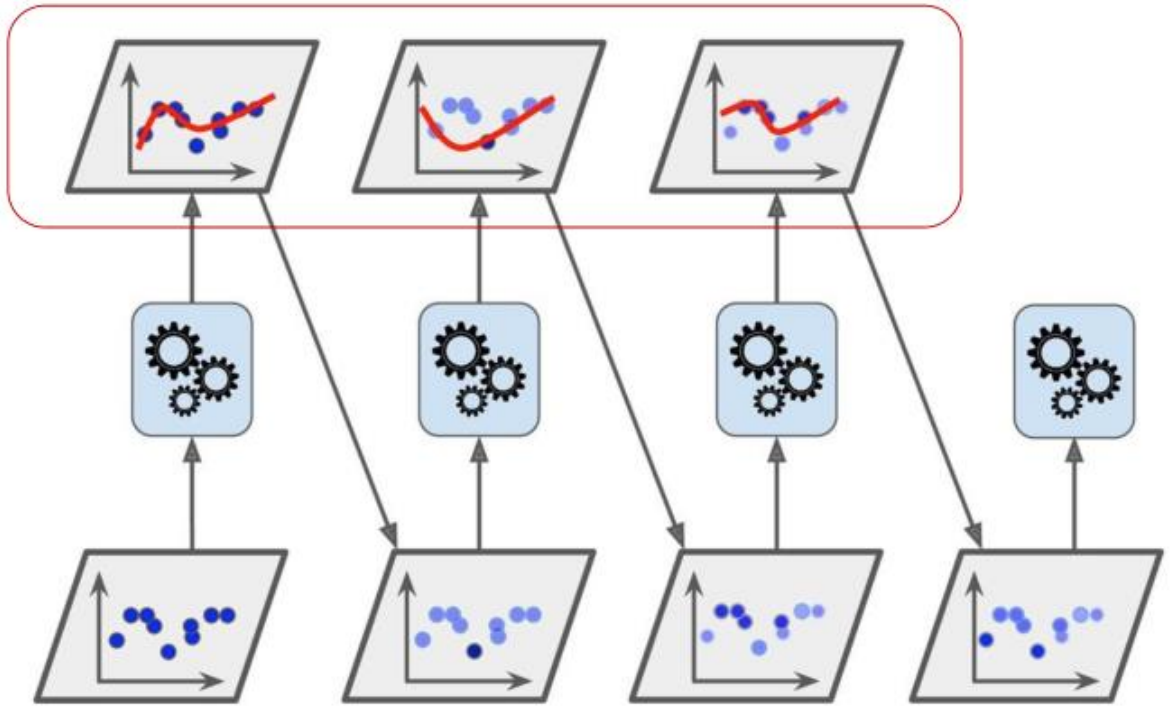
4. XGBoost is one efficient and scalable implementation of GBDT

Q4

4. CART are usually used as the base predictors of gradient-boosted decision trees (GBDT). Which of the following are correct?

1. GBDT constructs an ensemble of trees in parallel -> False (sequential not parallel)

2. During GBDT's fitting, a new CART predictor is trained using the residual from the last CART as the weight, considering the largest residuals -> True



Q4

Open Source Collective sponsors

backers 25 sponsors 8

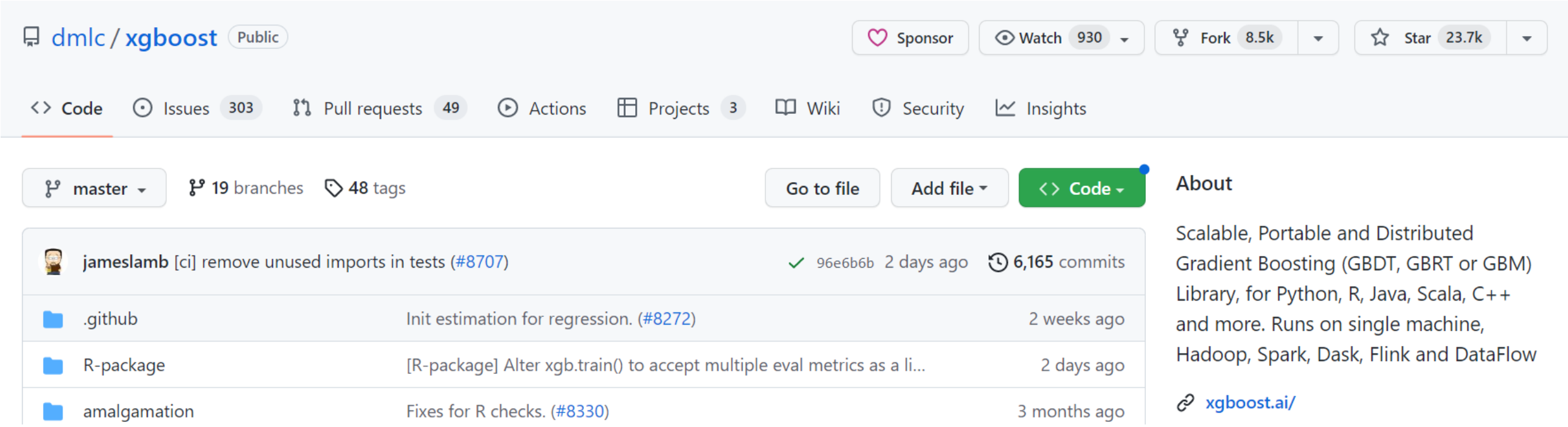
Sponsors

[Become a sponsor]

4. CART are usually used as the base predictors of gradient-boosted decision trees (GBDT). Which of the following are correct?

3. GBDT has been used successfully in many data science competitions -> True

4. XGBoost is one efficient and scalable implementation of GBDT -> True



Q5

Remember: CART are cheap to train and make no assumptions about the data -> good candidates for ensemble learning!

5. CART are usually used in the context of random forest (RF) and gradient-boosted decision trees (GBDT). Which of the following are correct?

1. RF and GBDT are ensemble learning methods based on CART which improve CART's lack of robustness (overfitting + smoothness)

2. RF and GBDT model non-linear relationships between features

3. Both RF and GBDT lack interpretability

4. By averaging multiple trees, RF's and GBDT's predictions are more robust compared to those of individual CARTs