Decision Trees

Benefits of Decision Trees

- Easy to interpret
- Easy to understand
- Extensions can add more to interpretability

But: Downside is that they are not very powerful for making good predictions.

Decision Trees

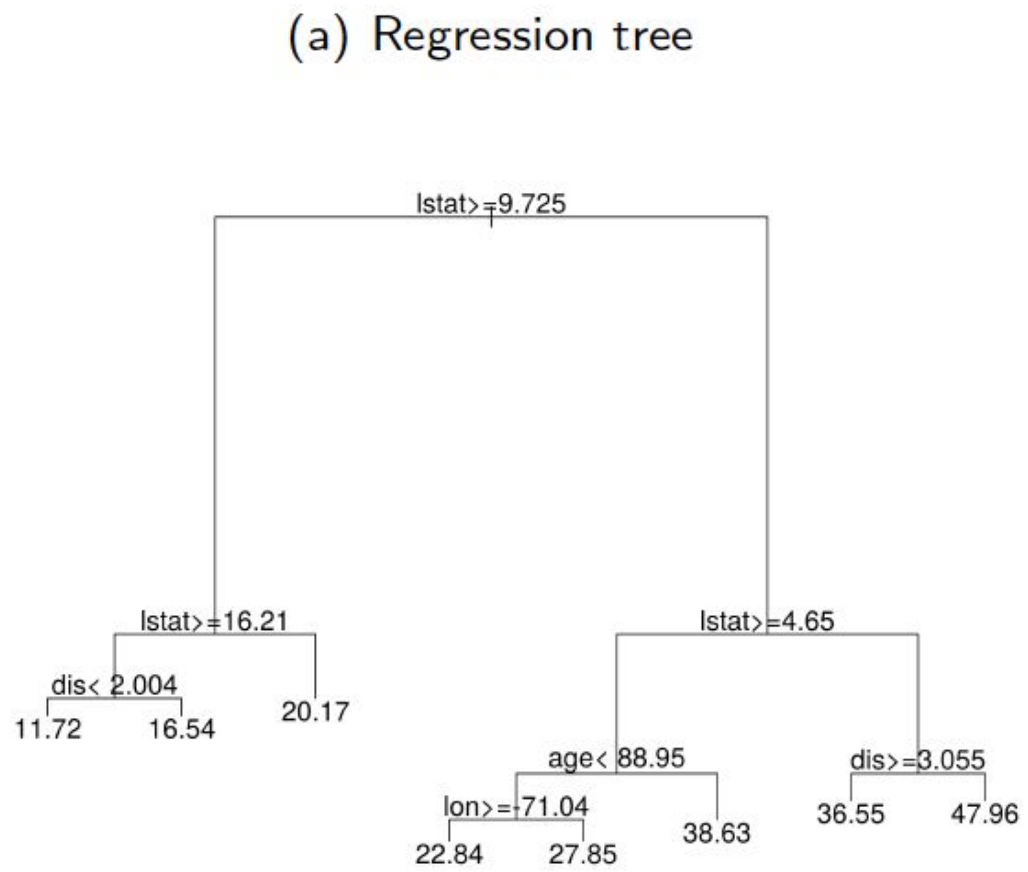
Note on terminology:

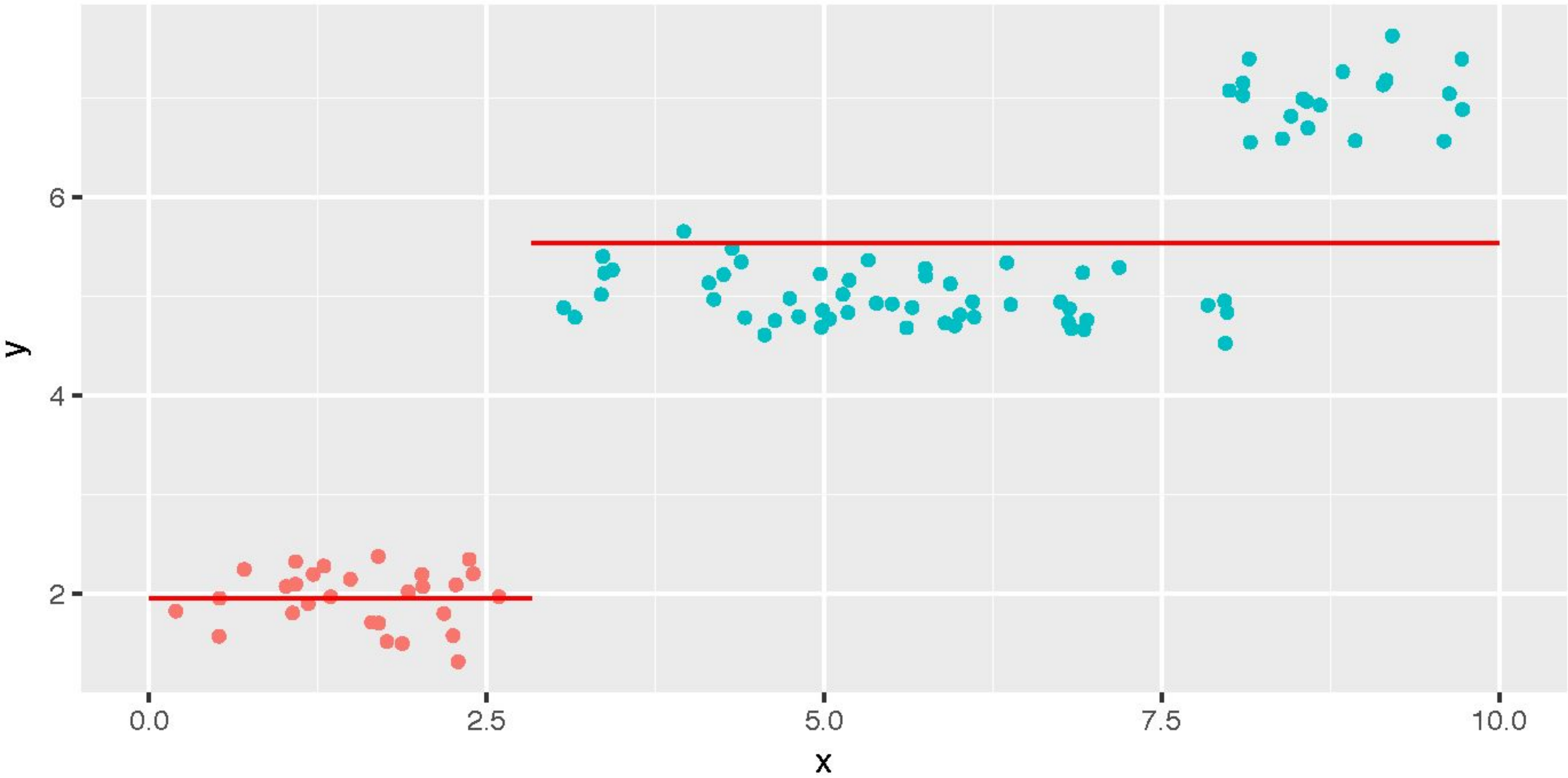
- Regression trees refer to trees used for regression (i.e., numerical outcome).
- Classification trees refer to trees used for classification (i.e., categorical outcome).

Interpreting Trees

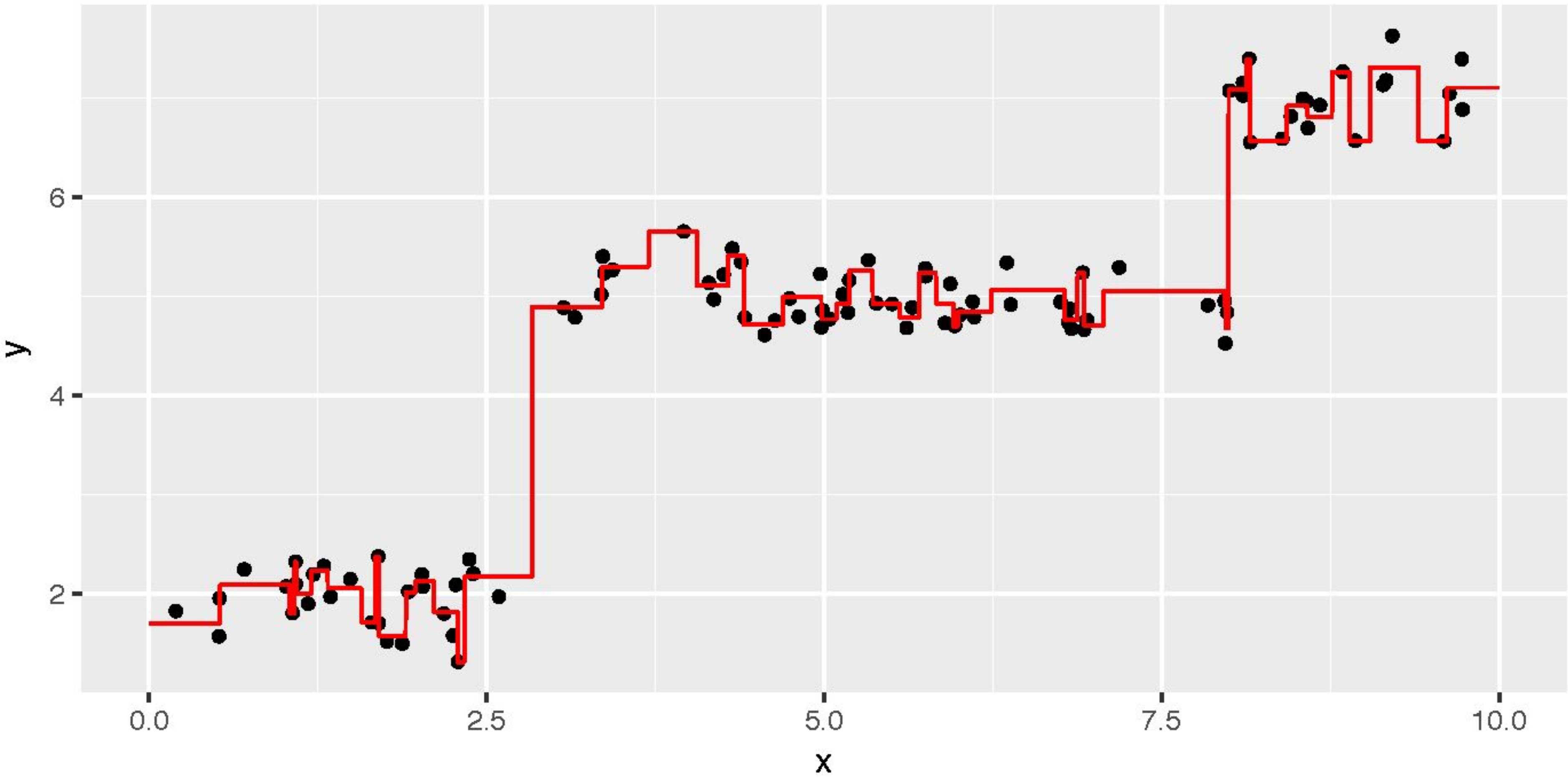
Follow the steps from the top to make a prediction for a given observation.

Terminal node shows estimate.

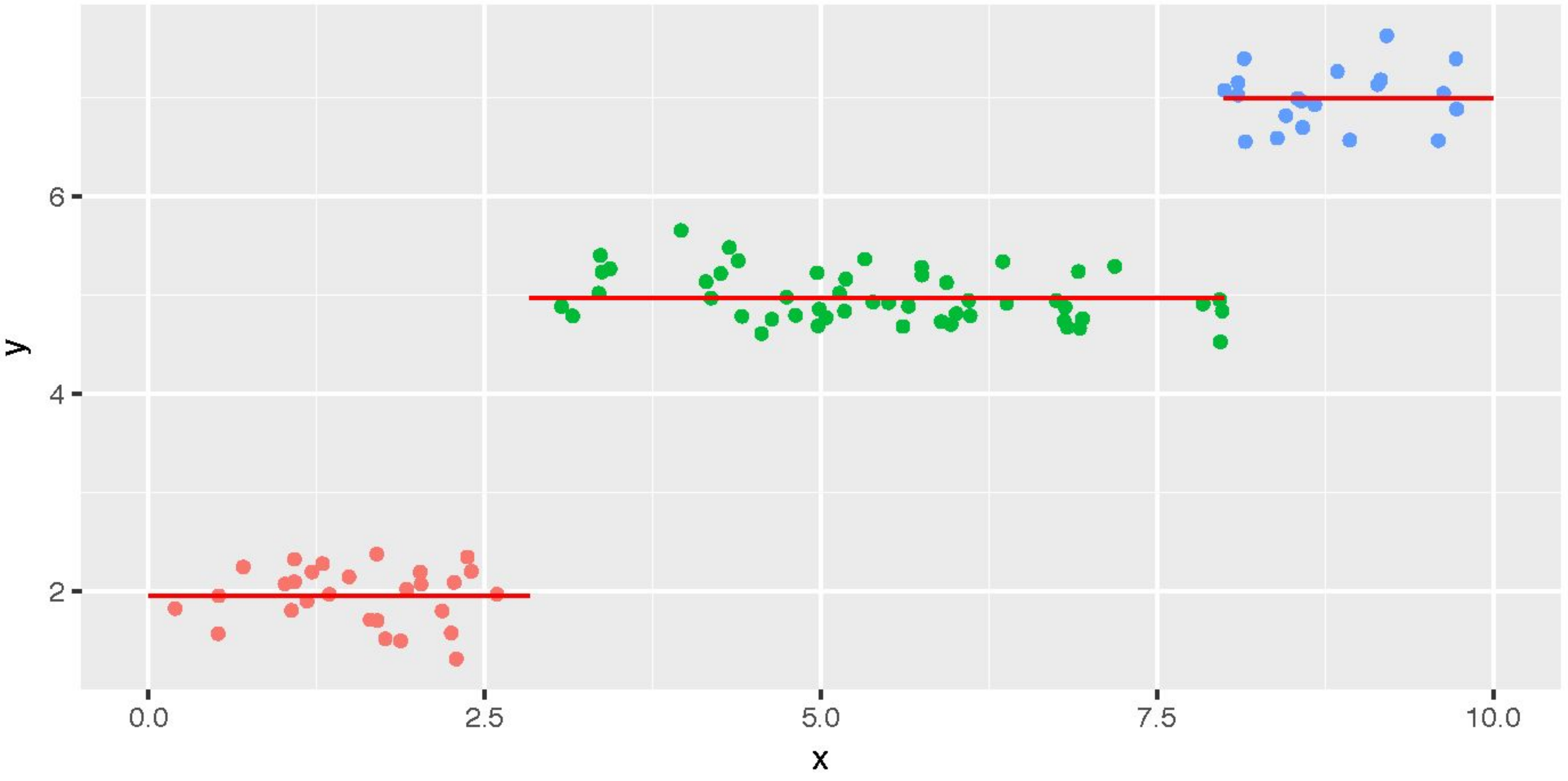




6
4
2
y
0.0
2.5
5.0
7.5
10.0
x



y
6
4
2
0.0
2.5
5.0
7.5
10.0
x



6
4
2
y
0.0
2.5
5.0
7.5
10.0
x

Dealing with Correlated Predictors

We'll see how we can use the idea of decision trees and deal with correlated predictors when we get to ensemble models

General idea: We only use a few of the predictors at a time.

Surrogate Splits

Used to address issues that can arise with missing data.

Idea: Determine split, then find another variable that mimics that split. Use the other variable if there is a missing value in the original variable.

Looking Ahead

Individual trees tend to **overfit.** How can we address this to get better models?

Ensemble methods such as Random Forests tries to address some of the issues with individual trees.

- Use only a subset of predictors at a time.
- Use a bootstrap sample of observations.

Variations on Trees

Conditional Inference trees and Model based recursive partitioning aim to put more of a structure on building trees. These provide more **interpretability**.

Generally, if prediction is our primary purpose, we’ll use **regular trees as base learners.**

Imbalanced Datasets

- Note: We **don’t need to set the majority class** of a terminal node as our prediction. We can get scores out of the proportion in the terminal node.
- Using **different metrics (and tuning)** can help identify the best model for prediction.
- There are methods to deal with imbalance data (we will discuss later).