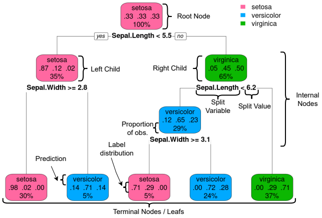
# Introduction to Machine Learning

# CART: In a nutshell




## Learning goals

- Understand basic structure of CART models
- Understand basic concepts used to fit CART models

# LEARNING AND PREDICTION WITH CARTS

## Training

Input: Labeled data

| Sepal Width | Sepal Length | Species |
|---|---|---|
| 3,5 | 5,1 | Setosa |
| 3,0 | 5,9 | Virginica |
| 2,5 | 6,3 | Virginica |
| 2,2 | 6,2 | Versicolor |

Tree Learner

Output

Tree Model

## Prediction

Input: Unlabeled data

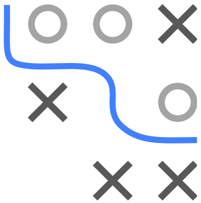
| Sepal Width | Sepal Length | Species |
|---|---|---|
| 3,0 | 7,7 | ? |
| 3,1 | 6,9 | ? |
| 2,3 | 6,3 | ? |
| 3,0 | 5,6 | ? |

Tree Model

Prediction

| Level of Happiness |
|---|
| Setosa |
| Versicolor |
| Virginica |
| Versicolor |

# WHAT IS A TREE?



Basic idea:

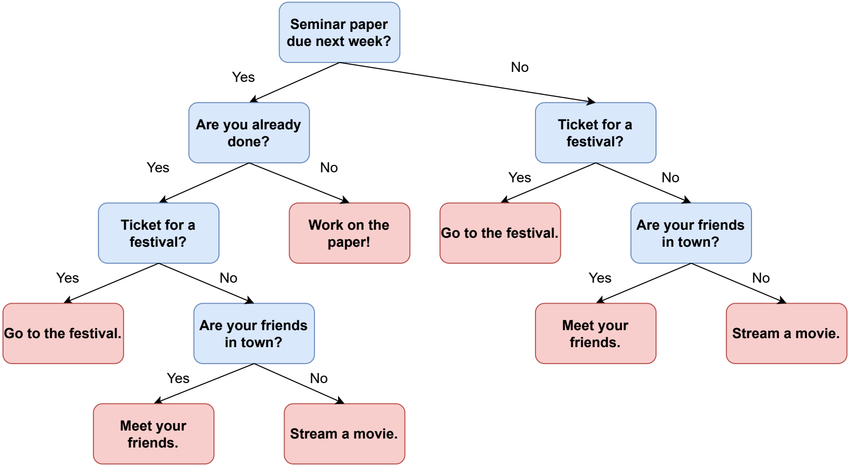
- Divide feature space into sub-regions.
- For each region, learn best constant prediction from training data.

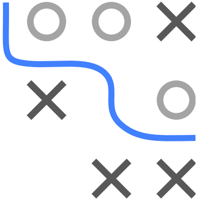
**C**lassification **A**nd **R**egression **T**rees are a class of models that can:

- model non-linear feature effects
- facilitate interactions of features
- be inherently interpreted

# WHAT IS A TREE?

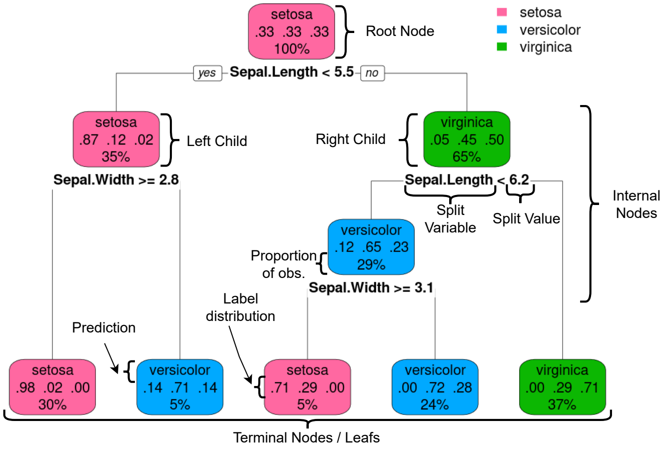
A decision tree is a set of hierarchical binary partitions, e.g., your evening planning decision (target) could be based on a decision tree:



# WHAT IS A TREE?

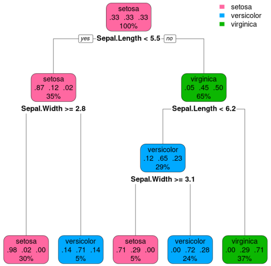
Instead of life choices, we can predict the `Species` of flowers described in the `iris` data set using features `Sepal.Width` and `Sepal.Length`.

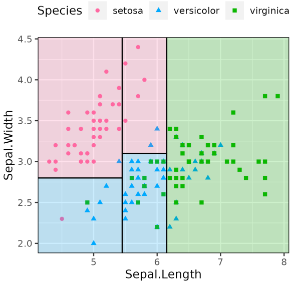


# CART AS A PREDICTOR

Instead of the visual description, we can also describe trees through their division of the feature space $\mathcal{X}$ into **rectangular regions**, $Q_m$:

$$f(\mathbf{x}) = \sum_{m=1}^{M} c_m \mathbb{I}(\mathbf{x} \in Q_m),$$

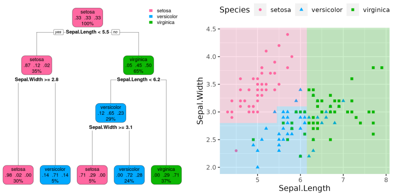


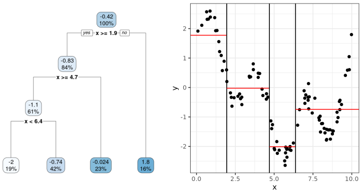
# TASKS FOR CART

- CARTs can have categorical and numerical targets.
- In both cases, the leafs, i.e., the ultimate nodes, define the predictions.

Categorical target:



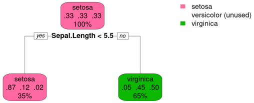

Numerical target:

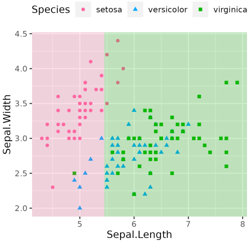


# HOW TO FIT A CART

- A *recursive* greedy search in the feature space optimizes CARTs
- In each iteration, the best split is selected

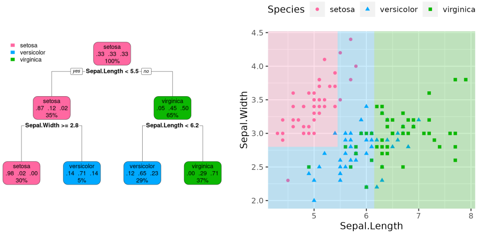
Iteration 1:



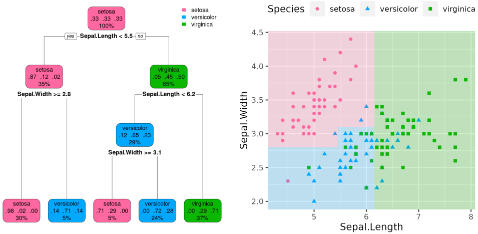



# HOW TO FIT A CART

Iteration 2:

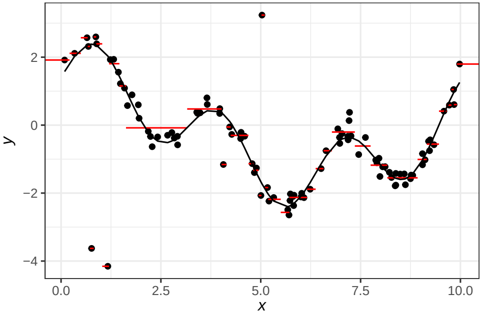



Iteration 3:



# HOW TO FIT A CART

- This procedure can run until each observation has its own leaf.
- Then, the tree will not generalize well and overfit:




- Thus, we need techniques to keep the tree small and informative.

⇒ In practice, trees are often used as base learners for ensemble learners like Random Forests.