Trees and Forests

William Brasic

The University of Arizona

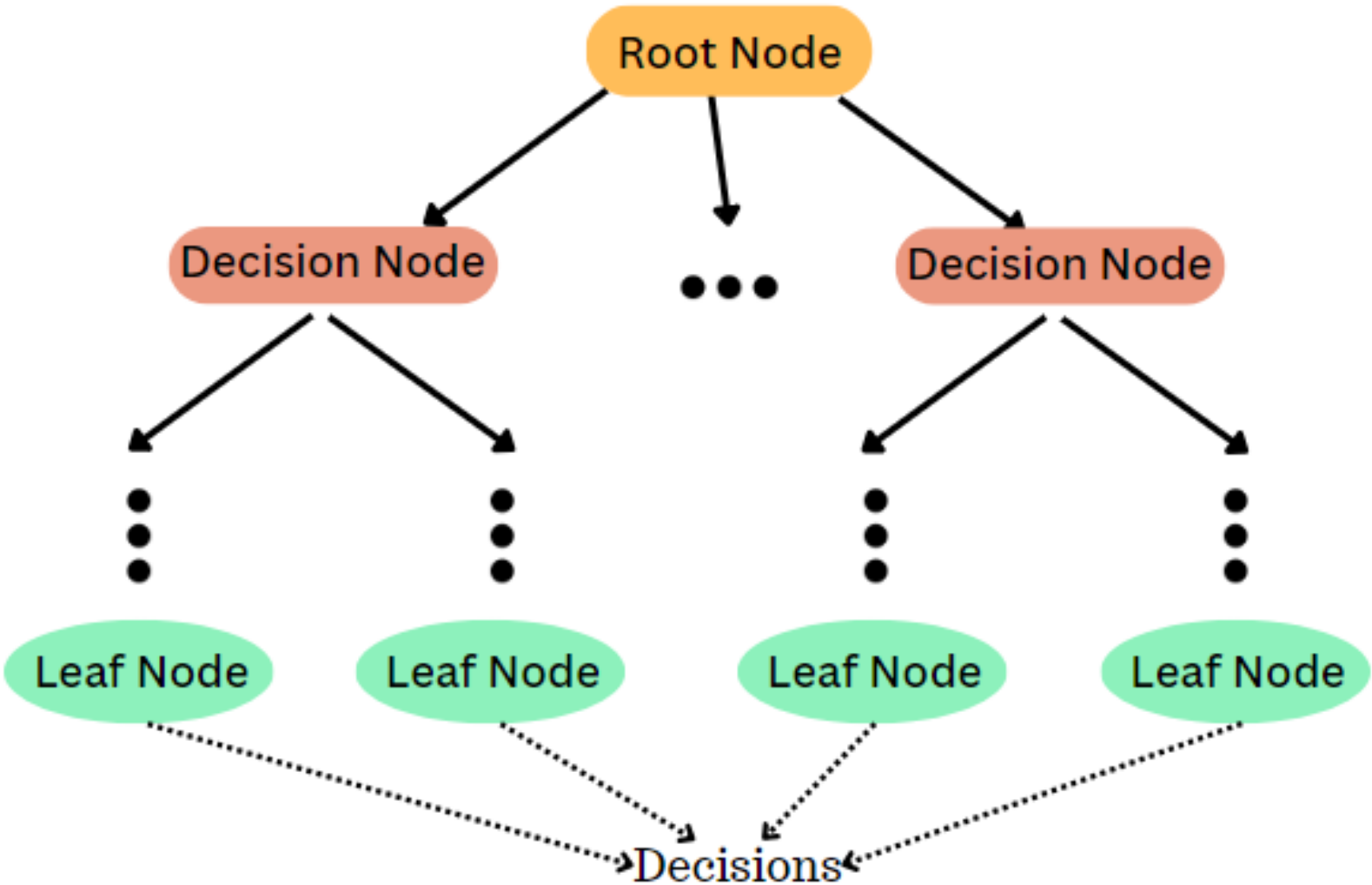
Decision Tree

Definition 1: Decision Tree

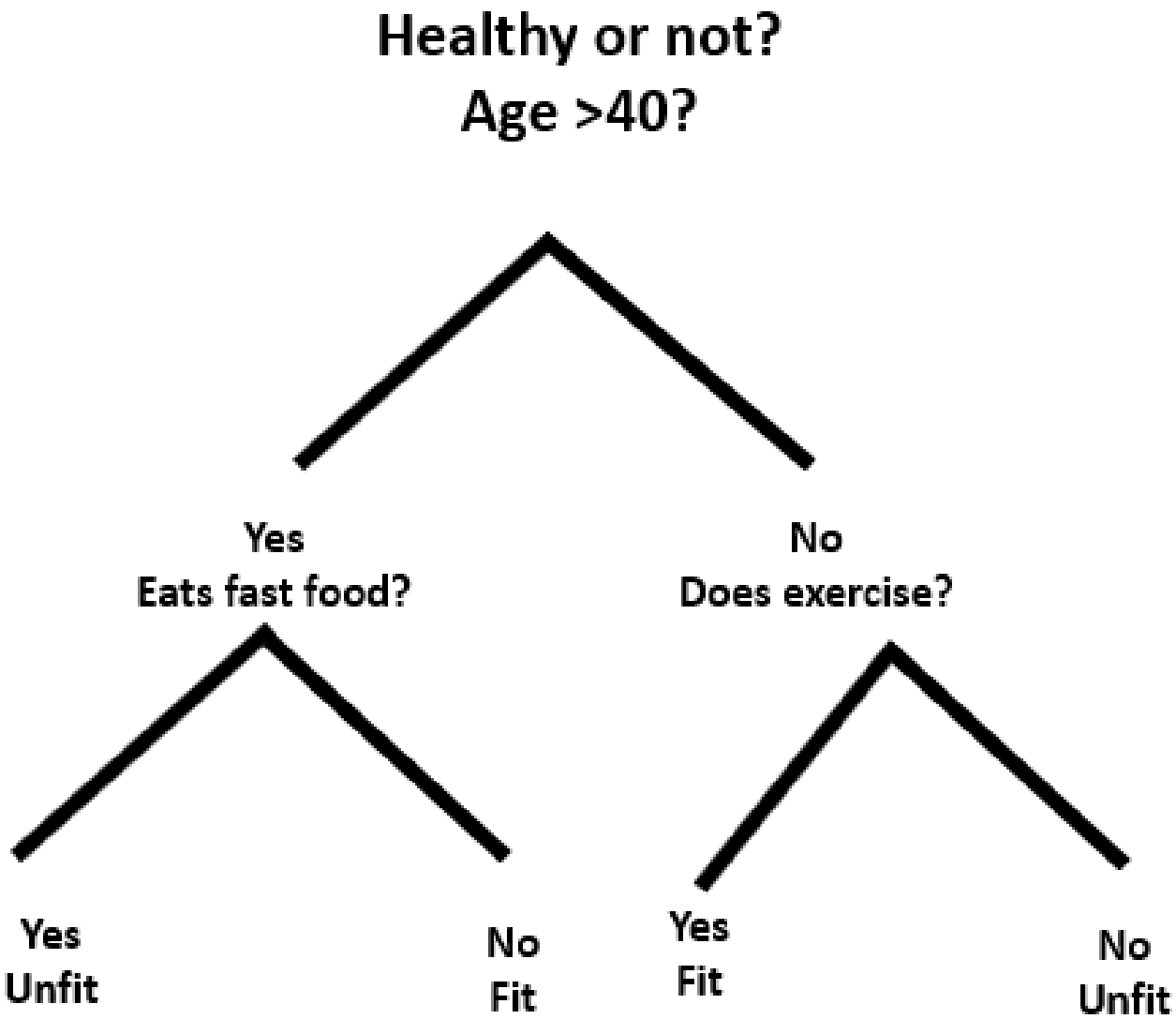
A decision tree is a non-parametric regression and classification algorithm that partitions the feature space into a series of rectangles and then fits a simple model.

- Decision trees for:
  - Regression are called *regression trees*.
  - Classification are called *classification trees*.
- A "simple model" is just the sample average.

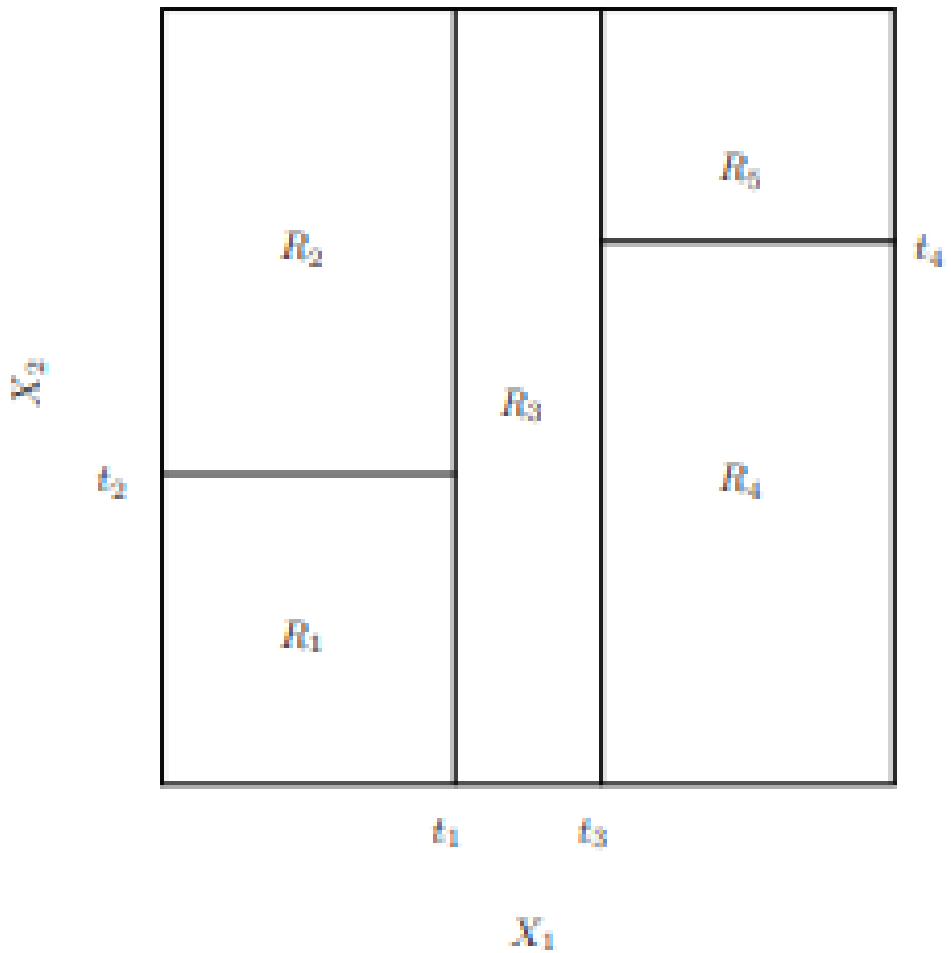
Decision Trees



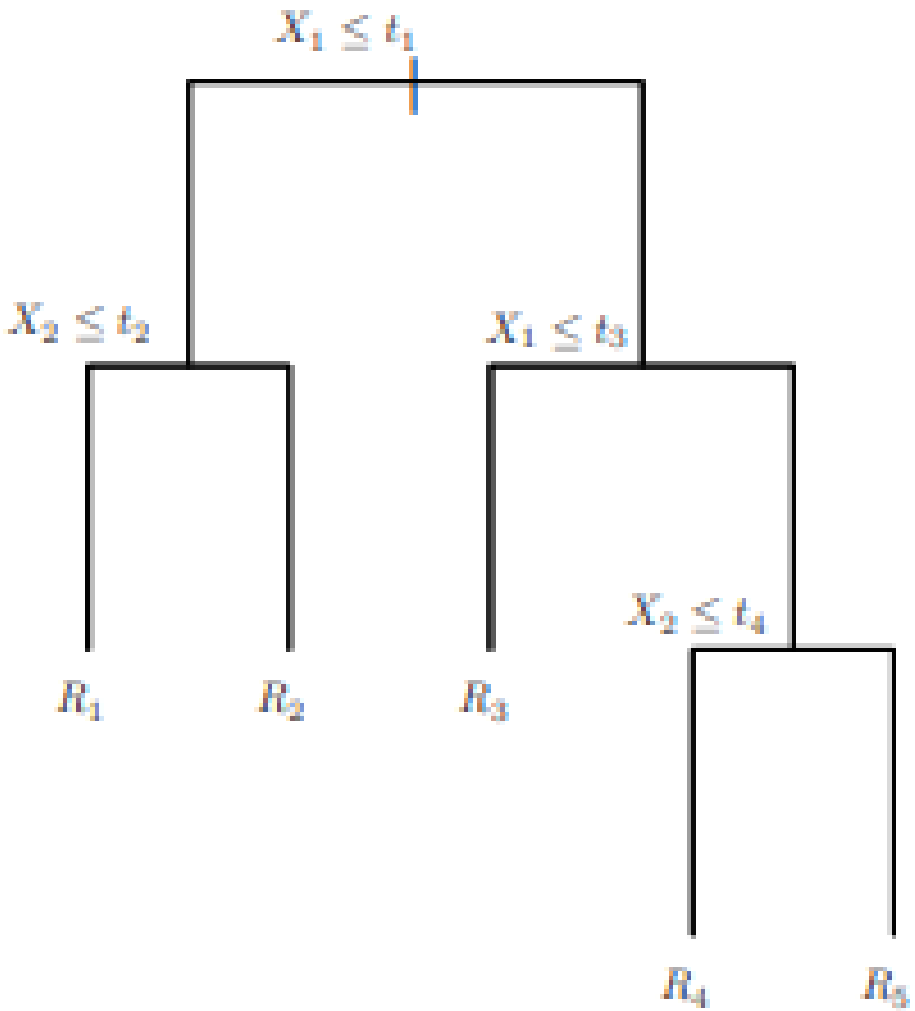
Decision Tree Example



Decision Tree Example



Decision Tree Example



Regression Tree Prediction

Definition 2: Regression Tree Prediction

The prediction using a regression tree t for a feature vector $\boldsymbol{x}$ is given by the average outcome across observations in the leaf node (region) R_m that $\boldsymbol{x}$ lands in:

$$\widehat{f}_t(\boldsymbol{x}) = \frac{1}{|R_m|} \sum_{i \in R_m} y_i.$$

Classification Tree Prediction

Definition 3: Classification Tree Prediction

The prediction using a classification tree t for a feature vector $\boldsymbol{x}$ is given by a majority vote across observations in the leaf node (region) R_m that $\boldsymbol{x}$ lands in:

$$\widehat{f}_t(\boldsymbol{x}) = \underset{c \in\{1,2,\ldots,C\}}{\arg\max} \left[\sum_{i \in R_m} \mathbb{1}(y_i = c) \right].$$

CART Algorithm

Question 1: CART Algorithm

How are the split point and split variables determined?

Answer to Question 1

The Classification and Regression Tree (CART) algorithm.

CART Algorithm

Definition 4: CART Algorithm

The CART algorithm follows these steps:

1. Start at the root node.

CART Algorithm

Definition 4: CART Algorithm

The CART algorithm follows these steps:

1. Start at the root node.
2. Select the best split (determine the best split point and best split variable).

CART Algorithm

Definition 4: CART Algorithm

The CART algorithm follows these steps:

1. Start at the root node.
2. Select the best split (determine the best split point and best split variable).
3. Partition the data based on the split.

CART Algorithm

Definition 4: CART Algorithm

The CART algorithm follows these steps:

1. Start at the root node.
2. Select the best split (determine the best split point and best split variable).
3. Partition the data based on the split.
4. Repeat steps 2 and 3 until some stopping criteria is met (such as depth of the tree).

CART Algorithm

Definition 4: CART Algorithm

The CART algorithm follows these steps:

1. Start at the root node.
2. Select the best split (determine the best split point and best split variable).
3. Partition the data based on the split.
4. Repeat steps 2 and 3 until some stopping criteria is met (such as depth of the tree).
5. Prune the tree.

- Tree depth is a hyperparameter that needs to be tuned.
- Tree pruning is a technique used to reduce the depth of the tree to avoid overfitting.

Cost Complexity Pruning

Definition 5: Cost Complexity Pruning

Cost complexity pruning is a technique used to prune a decision tree to avoid overfitting by considering a trade-off between the complexity of the tree and its fit to the training data.

- *Pruning* is the process by which we try to remove unnecessary leaf nodes from the original tree to reduce overfitting.
- The goal is to find the subtree T_α of the original T that generalizes better.
  - α penalizes decision tree for being too deep (similar to λ in ridge and lasso regression).

Decision Trees Pros

Property 1: Decision Trees Pros

1. Simple to understand and interpret.
2. Requires little data preprocessing (no need for normalization or scaling).
3. Can handle both numerical and categorical data.
4. Capable of capturing non-linear relationships.

Decision Trees Cons

Property 2: Decision Trees Cons

1. Prone to overfitting (high variance).
2. Can be unstable (small changes in data can lead to different trees).

Random Forest

Definition 6: Random Forest

A random forest is a non-parametric regression and classification algorithm that combines the predictions of T decision trees to form a final prediction.

- This final prediction is the sample average of predictions produced by the T trees for regression or the majority vote for classification.
- A random forest is an example of *bagging* (bootstrap aggregating) *ensemble learning* method.

The "Random" in Random Forests

Question 2: The "Random" in Random Forests

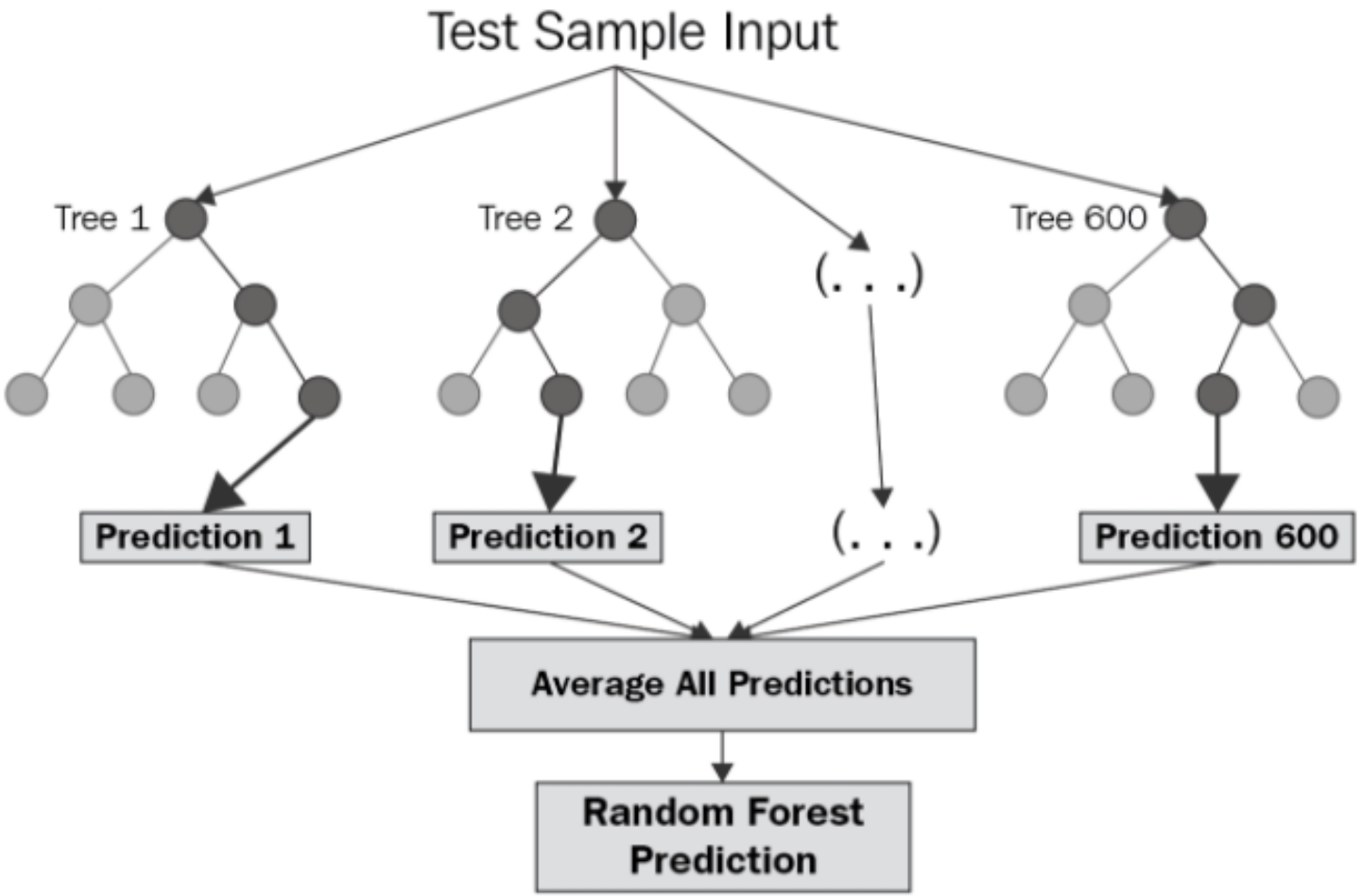
Why are random forests *"random"*?

Answer to Question 2

1. Each decision tree $t = 1, \ldots, T$ is fit using a different *bootstrapped* sample.
2. Each decision tree uses only a subset of the total possible features for splitting.

- *Bagging* greatly helps to reduce the high variance of learning algorithms like decision trees.

What is a Random Forest?



Random Forest Regression Prediction

Definition 13: Random Forest Regression Prediction

The prediction using a random forest regression for a feature vector $\boldsymbol{x}$ is given by the average prediction produced by the T trees in the random forest:

$$\widehat{f}(\boldsymbol{x}) = \frac{1}{T}\sum_{t=1}^{T}\widehat{f}_t(\boldsymbol{x}).$$

Random Forest Classification Prediction

Definition 14: Random Forest Classification Prediction

The prediction using a classification random forest for a feature vector $\boldsymbol{x}$ is given by a majority vote across the T trees in the random forest:

$$\widehat{f}(\boldsymbol{x}) = \underset{c \in \{1,\ldots,C\}}{\arg\max} \sum_{t=1}^{T} \mathbb{1}\left(\widehat{f}_t(\boldsymbol{x}) = c\right).$$

Thank You!