Implementing Ensemble Learning Using Averaging Methods

Janani Ravi
CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Bootstrap Aggregation i.e. bagging

Pasting vs. bagging

Random forests as ensembles of decision trees

Random patches

Random subspaces

Build more random, diverse individual predictors using Extra Trees

Individual learners in an ensemble are **independent** of other learners

Individual learners can be trained in **parallel**

Bagging and Pasting

Important Questions in Ensemble Learning

What kind of individual learners to use?

How should individual learners be trained?

How should individual learners be combined?

“If everyone in the room is thinking the same thing, then somebody isn’t thinking.”

General Patton

Bagging and Pasting

An ensemble (collection) of models, in which individual models are trained on different random subsets of training data.

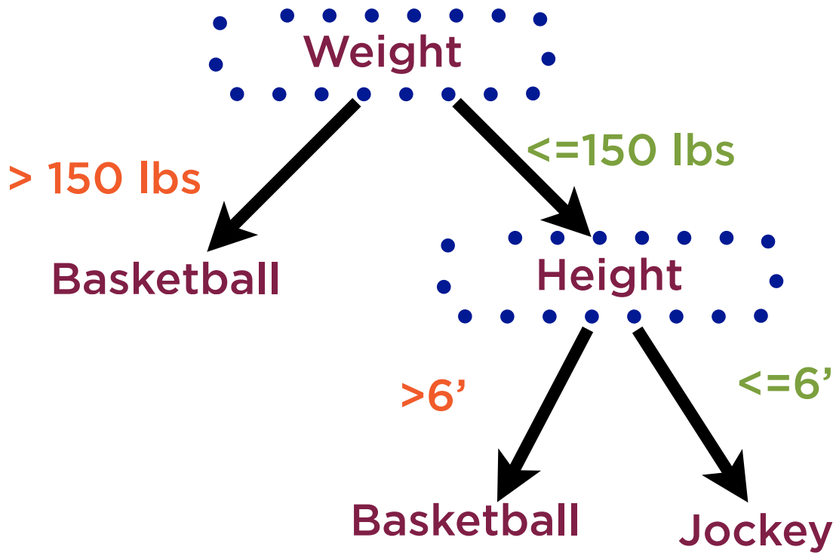
Ensemble Learning

Individual learners should be as different as possible

For most techniques, hard to generate large number of very different models

To the rescue: Decision trees and random forests

Random Forests



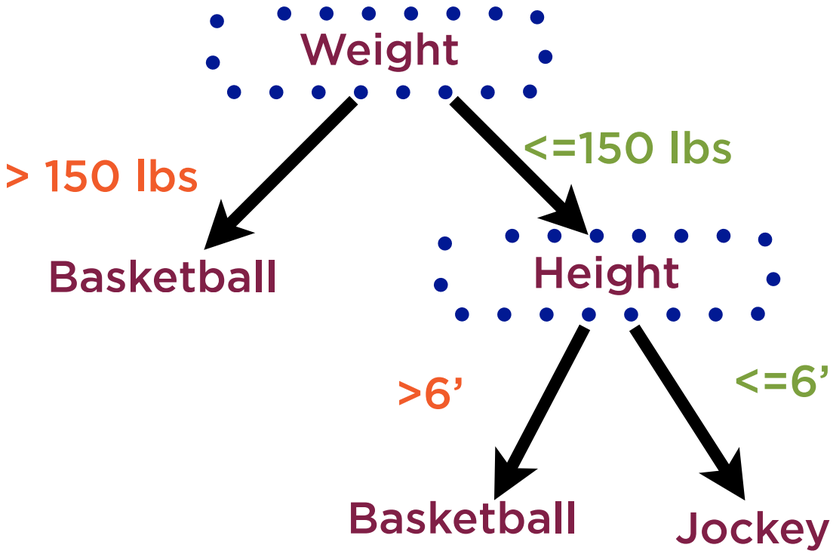
Train many decision trees

- each on random sample of data
- random set of features at split points

Combine their output

- averaging for regression
- mode for classification

Random Forests

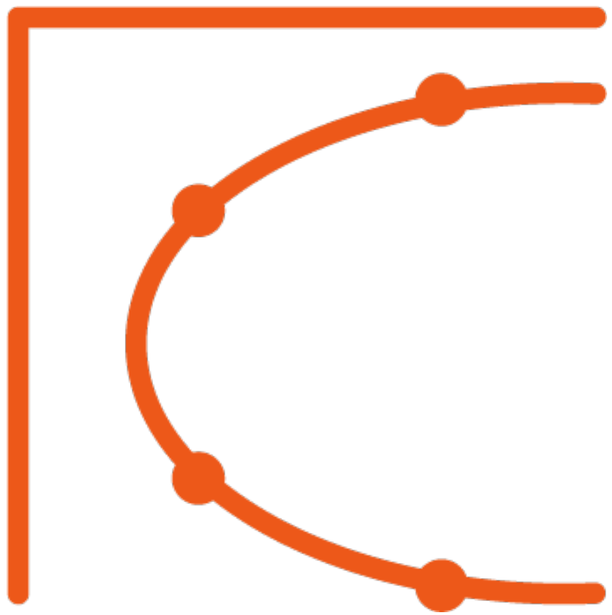


Extremely powerful technique

Example of ensemble learning

Individual trees should be as different as possible

Training Individual Decision Trees



Each iteration of training individual decision tree

- Sampling with replacement: Bagging or Bootstrap Aggregation
- Sampling without replacement: Pasting

Training Individual Decision Trees

Note that replacement refers to iterative training of individual trees

- Same point can still be used for training different trees in forest

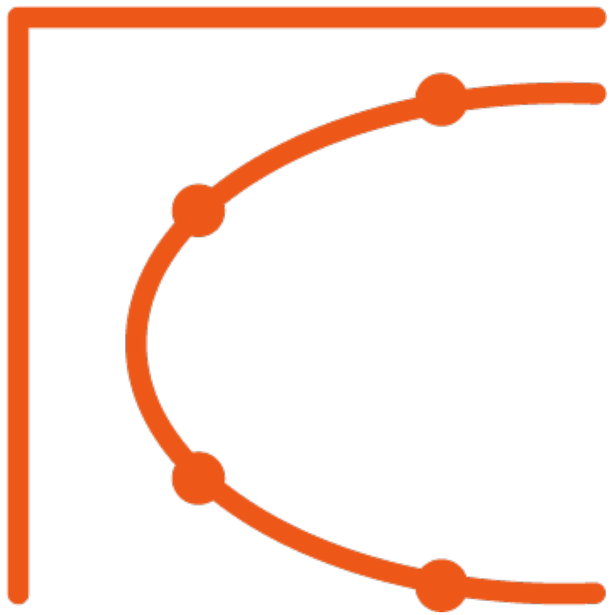
Training Individual Decision Trees

Bagging is usually preferred

- Tends to yield more uncorrelated predictors
- Less overfitting (lower variance error)
- Higher bias error than pasting

Averaging methods like bagging and pasting **scale very well** since models are trained in parallel

scikit-learn Hyperparameters



Bagging: bootstrap = True

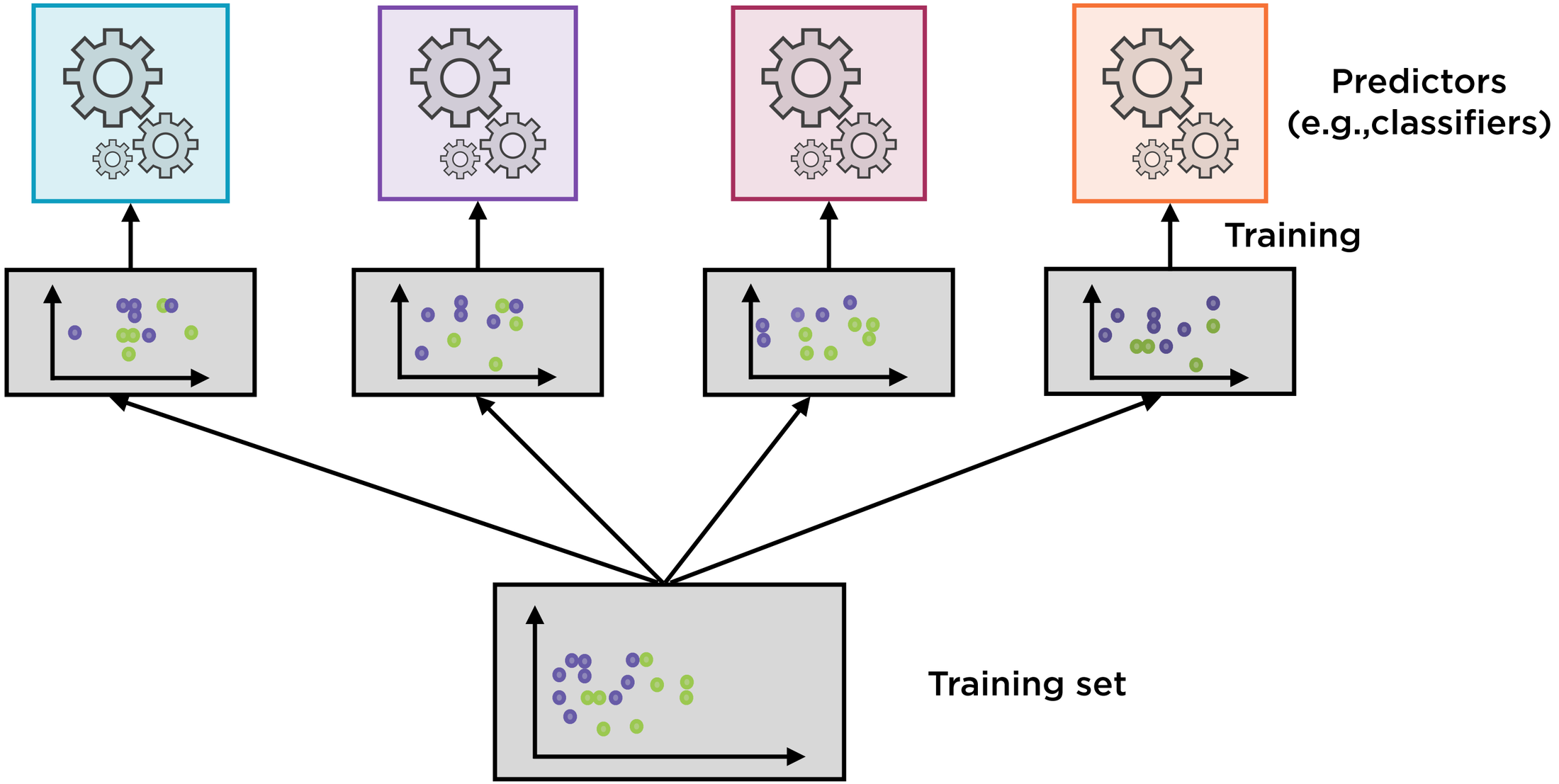
Pasting: bootstrap = False

scikit-learn Hyperparameters

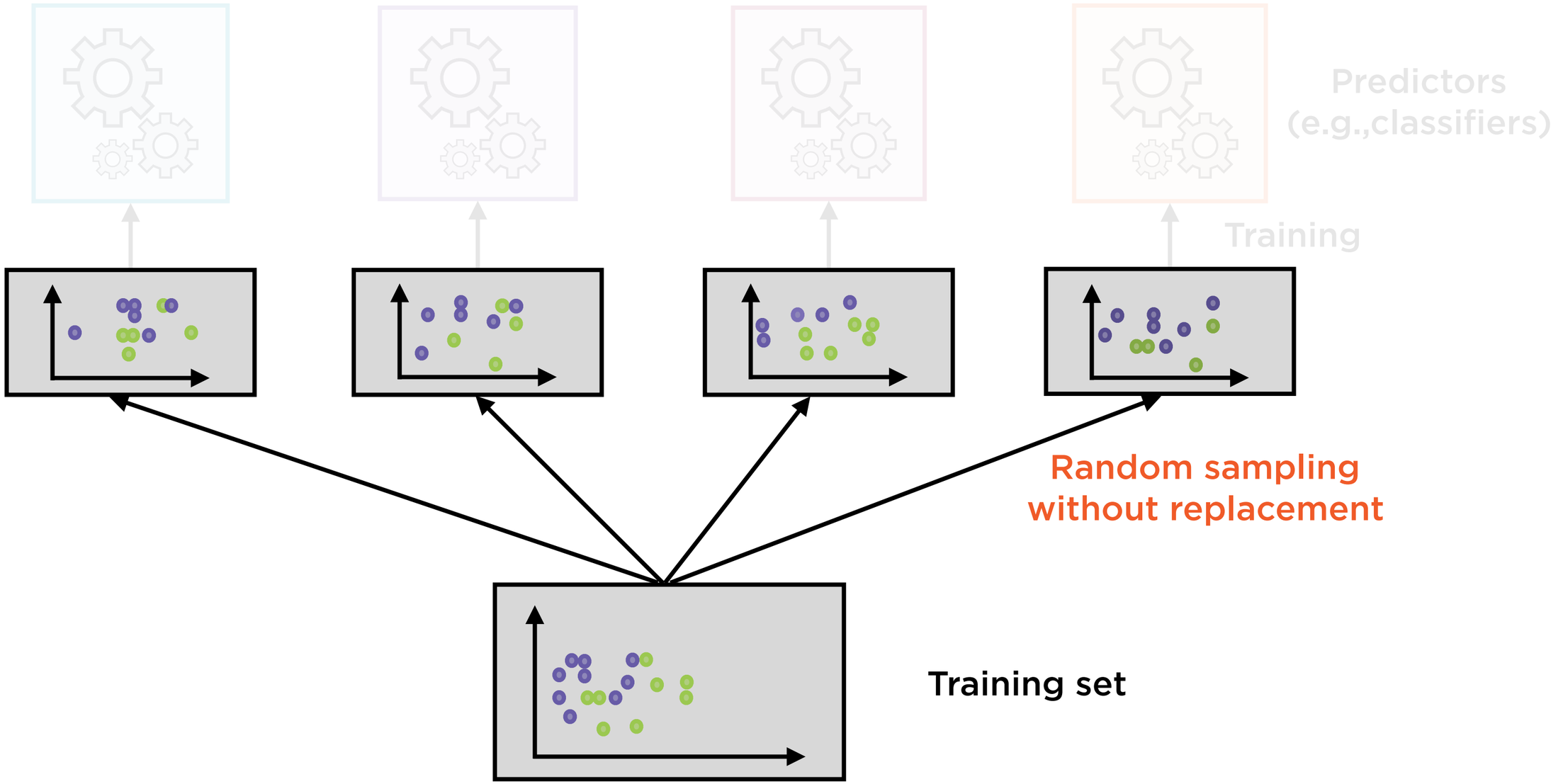
oob_score = True

- Out-of-bag evaluation
- Evaluate ensemble on those training points that are not seen in training

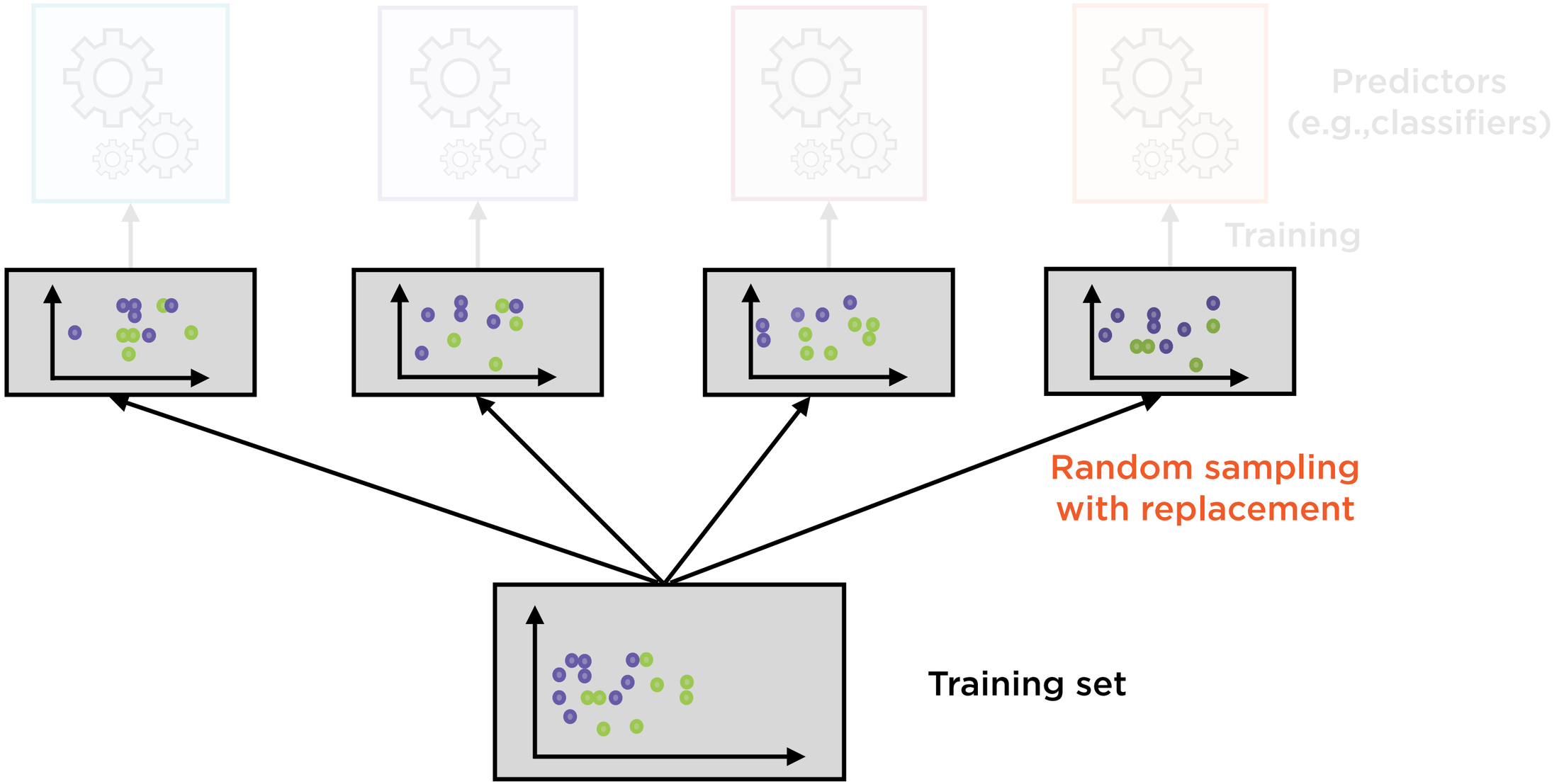
Sampling Data



Pasting - No Replacement



Bagging - With Replacement



Random Subspaces and Random Patches

Bagging and Pasting

An ensemble (collection) of models, in which individual trees are trained on different random subsets of training data.

Random Forest

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of training data.

Sampling Training Data

Age	Gender	Designation	Experience

Sampling Training Data

Age	Gender	Designation	Experience

Sampling Training Data

Age	Gender	Designation	Experience

Random Forest

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of training data.

Random Subspace

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of features in the training data but keeping all points in the training data.

Sampling Training Features

Age	Gender	Designation	Experience

Sampling Training Features

Age	Gender	Designation	Experience

Random Subspace

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of features in the training data but keeping all points in the training data.

Random Patches

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of features in the training data as well as random subsets of training data points.

Sampling Training Data and Features

Age	Gender	Designation	Experience

Sampling Training Data and Features

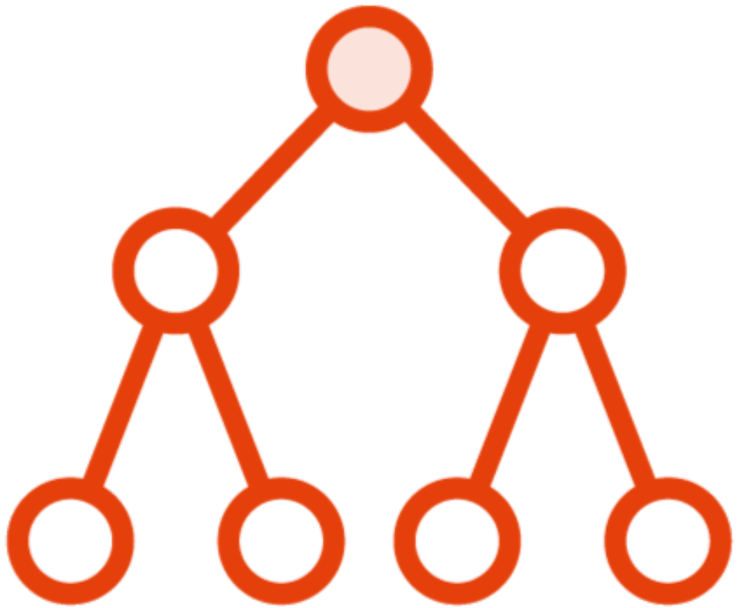
Age	Gender	Designation	Experience

Random Patches

An ensemble (collection) of decision trees, in which individual trees are trained on different random subsets of features in the training data as well as random subsets of training data points.

Extra Trees

Random Forest



Ensemble of decision trees

Usually assembled using bagging

During training

- Random subset of data
- Random subset of features at each split point

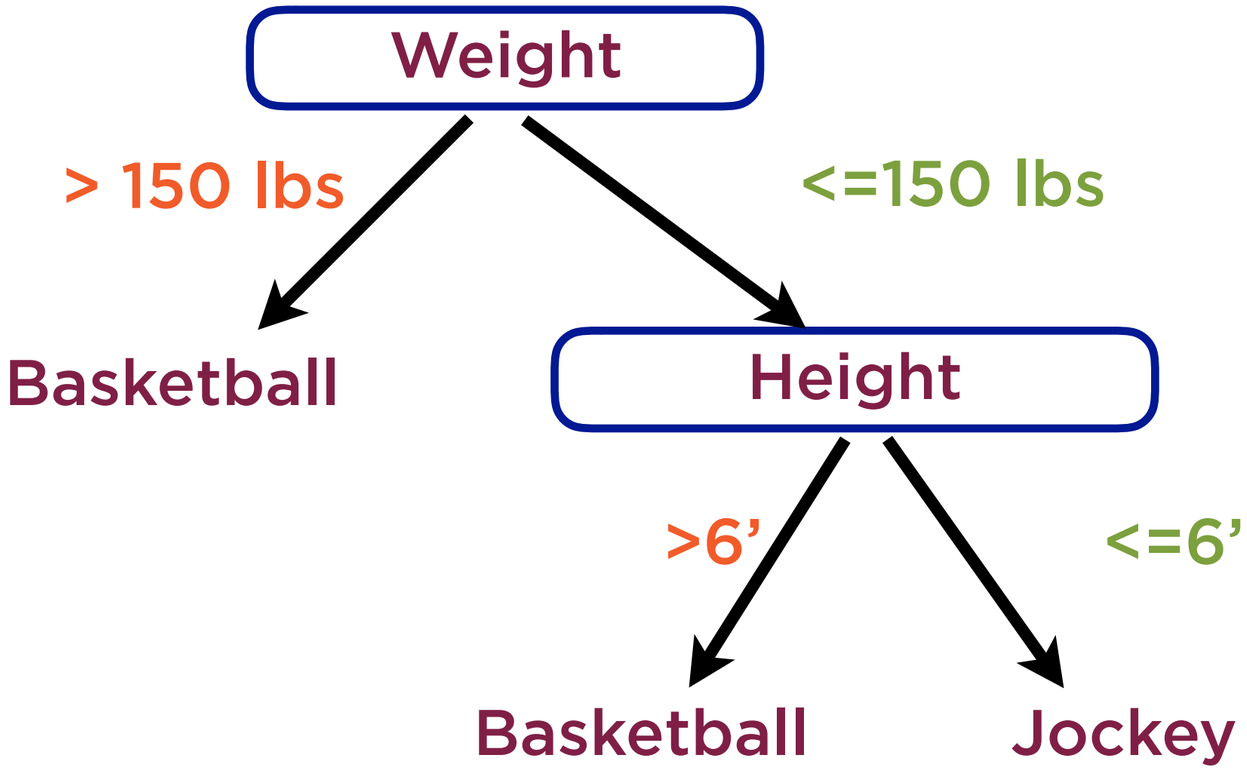
*Ext*remely *Ra*ndomized (Extra) Trees

An ensemble of decision trees, in which individual trees are trained on different random subsets of features in the training data using random split points (rather than finding best thresholds).

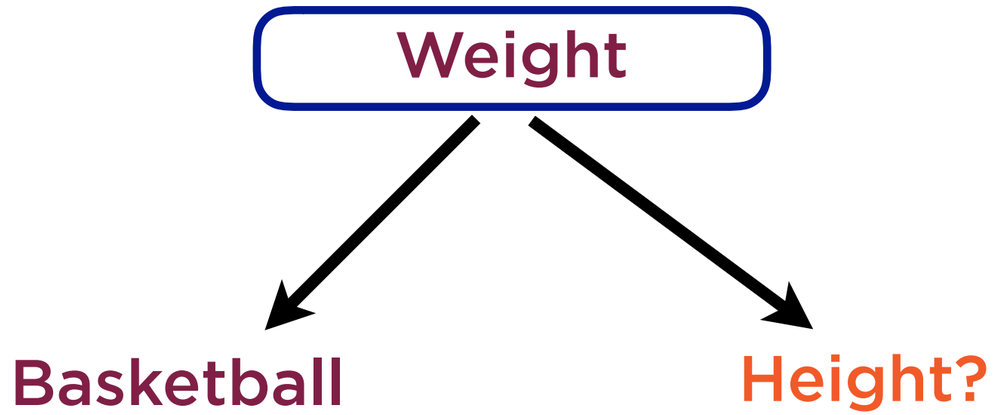
*Ext*remely *Ra*ndomized (Extra) Trees

An ensemble of decision trees, in which individual trees are trained on different random subsets of features in the training data using random split points (rather than finding best thresholds).

Fit Knowledge into Rules

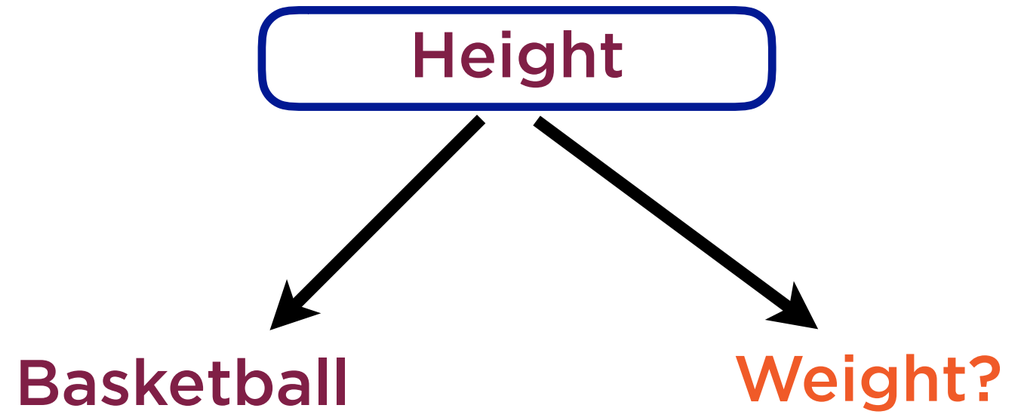


Decision Based on Weight



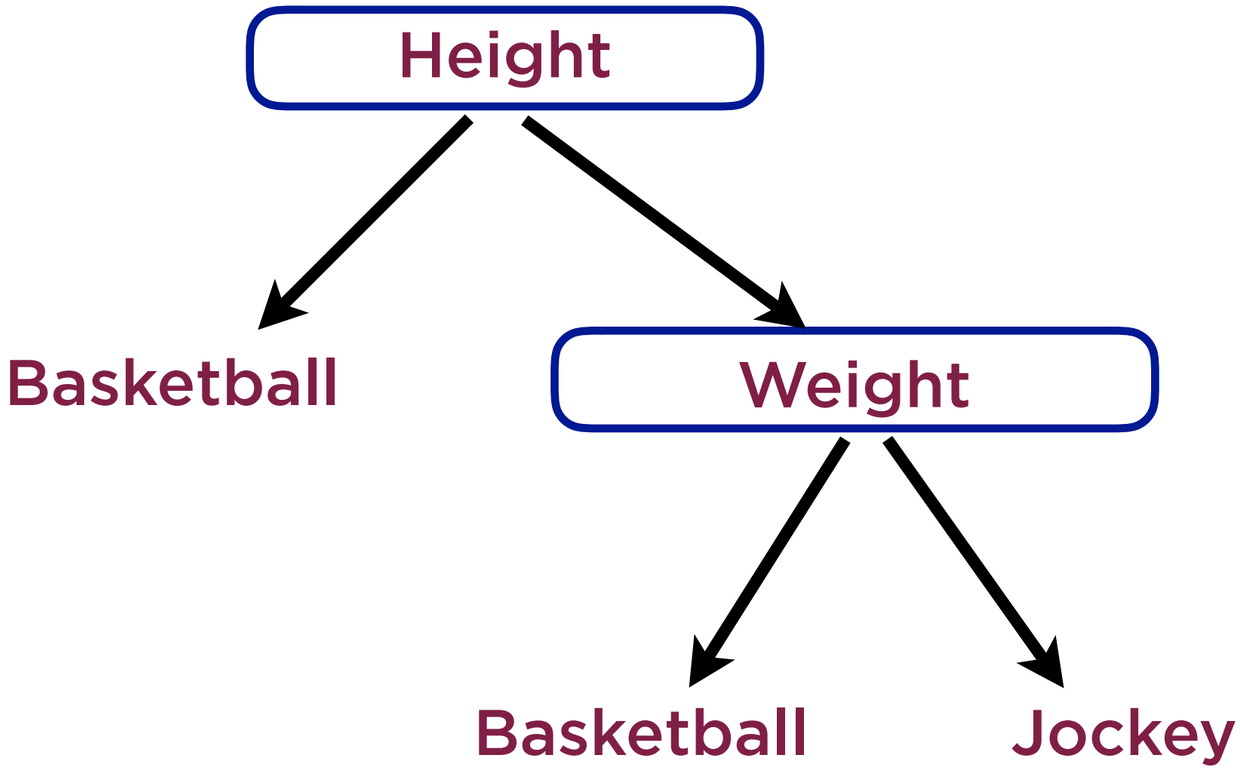
Choose the feature to split on at random - don't look for the best

Decision Based on Height

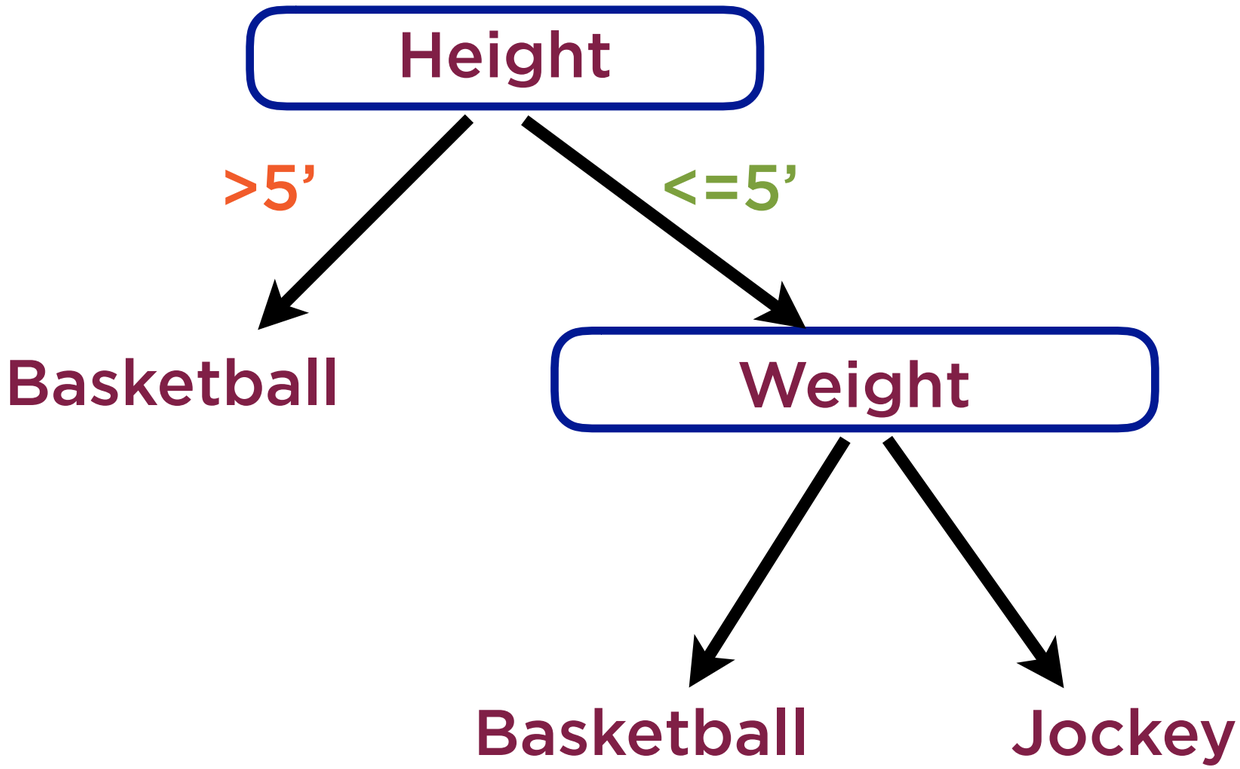


Choose the feature to split on at random - don't look for the best

Diverse Trees

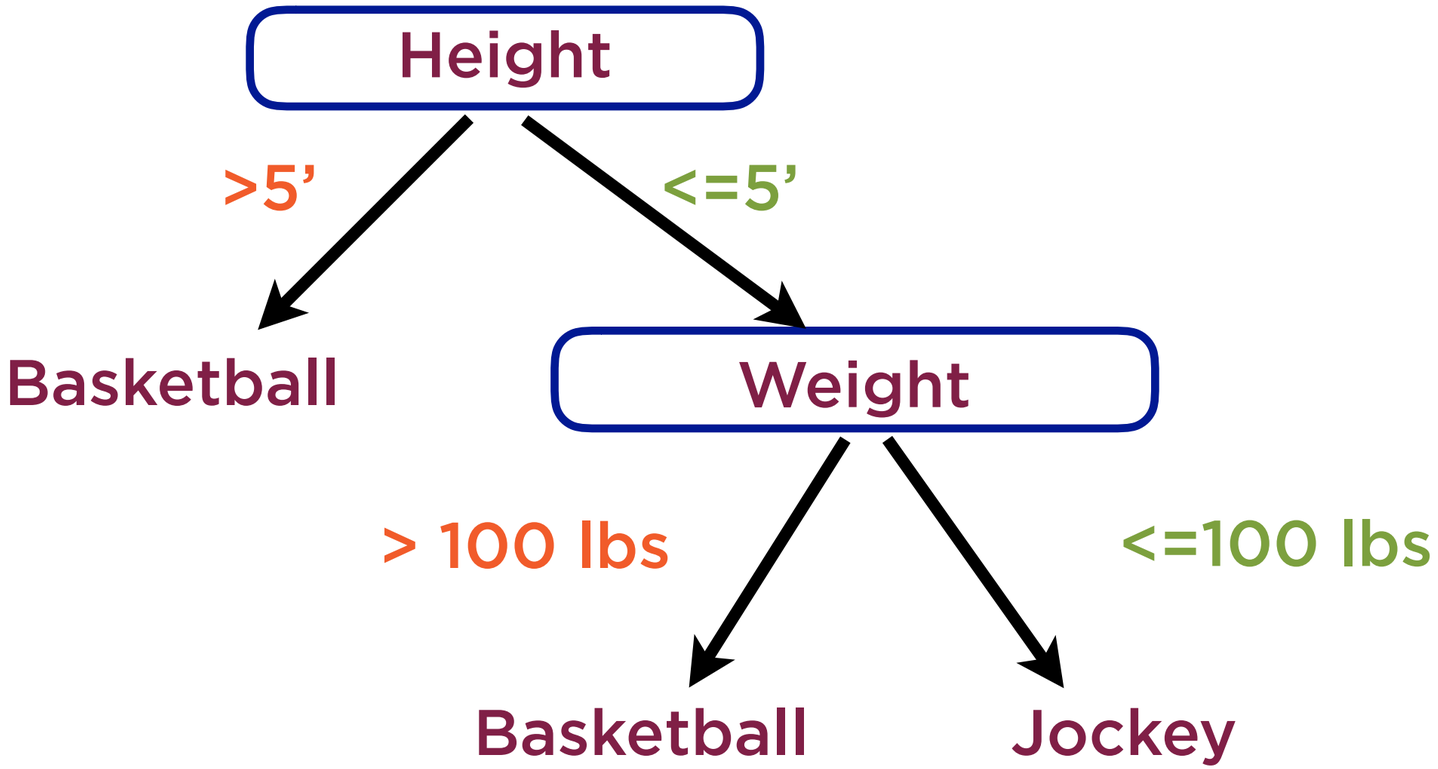


Fit Knowledge into Rules



Choose the threshold to split on at random - don’t look for the best

Fit Knowledge into Rules



Choose the threshold to split on at random - don’t look for the best

*Ext*remely *Ra*ndomized (Extra) Trees

An ensemble of decision trees, in which individual trees are trained on different random subsets of features in the training data using random split points (rather than finding best thresholds).

Averaging vs. Boosting

Important Questions in Ensemble Learning

What kind of individual learners to use?

How should individual learners be trained?

How should individual learners be combined?

So far - each individual learner was independent of all others - averaging methods

Averaging vs. Boosting

Averaging

Individual learners are independent

Can build trees in parallel

Learners do not learn from mistakes of other learners

Boosting

Individual learners are linked to previous learners

Need to build tree sequentially

Individual learners explicitly configured to learn from previous mistakes

Demo

Building ensemble regression models using bagging and pasting

Demo

Building ensemble classification models using bagging and pasting

Demo

Performing regression using the random forest and extra trees ensemble techniques

Demo

Performing classification using the random forest and extra trees ensemble techniques

Summary

Bootstrap Aggregation i.e. bagging

Pasting vs. bagging

Random forests as ensembles of decision trees

Random patches

Random subspaces

Build more random, diverse individual predictors using Extra Trees