Machine Learning with XGBoost Using Scikit-learn in Python

INTRODUCING ESSENTIAL PROCESSES

Mike West
MACHINE LEARNING ENGINEER

Module Overview

Examining the decision tree

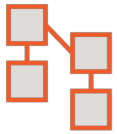
Ensemble boosting

Gradient boosting

Skills required for the course

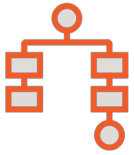
Building an XGBoost model

Why Take This Course?



One of the most accurate models in the world is XGBoost

The most winning model in competitive modeling is XGBoost



XGBoost excels when using structured datasets

Course Overview

Detailing the fundamentals

Preparing data for modeling

Scoring the XGBoost models

Saving the completed model

Feature selection

Summary

Your Skills

Not Required

Programming

Advanced machine learning

Statistics or advanced math

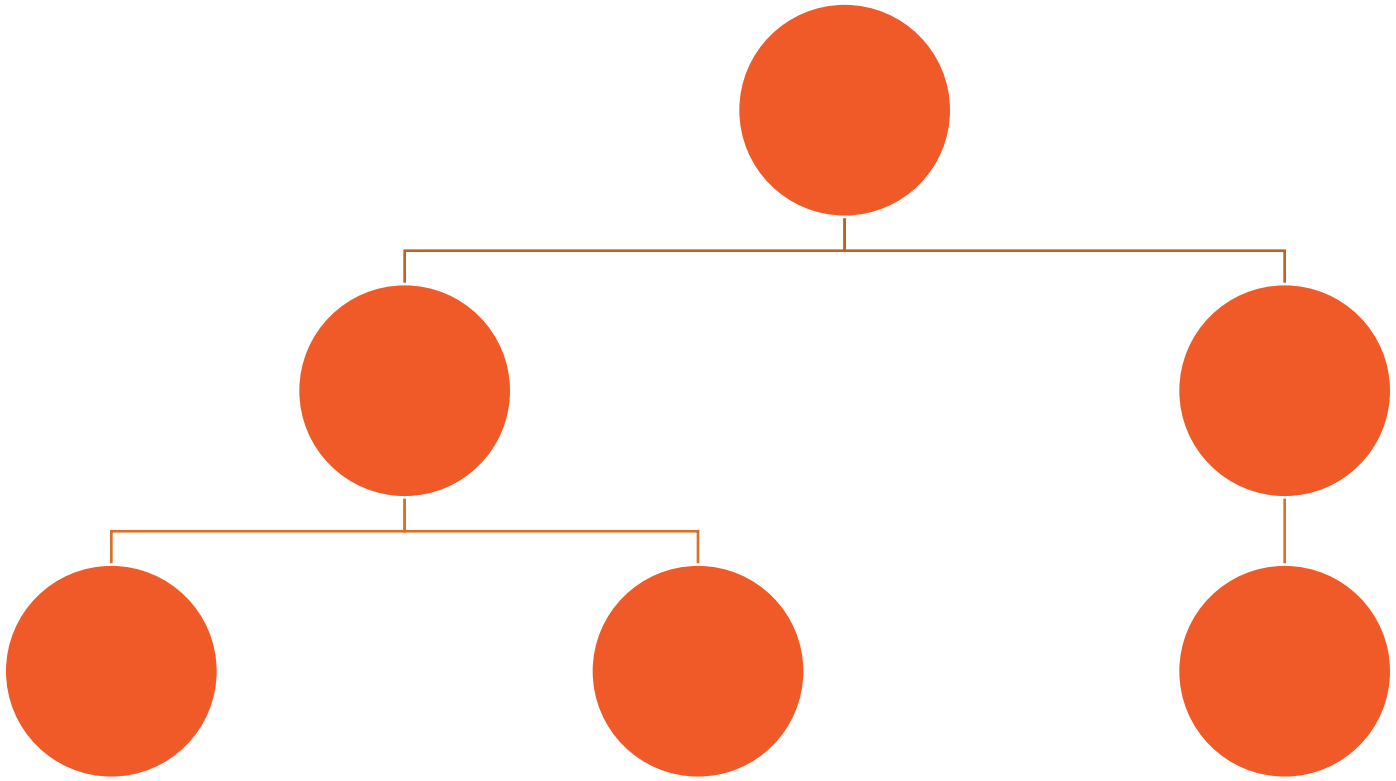
Required

Basic Python

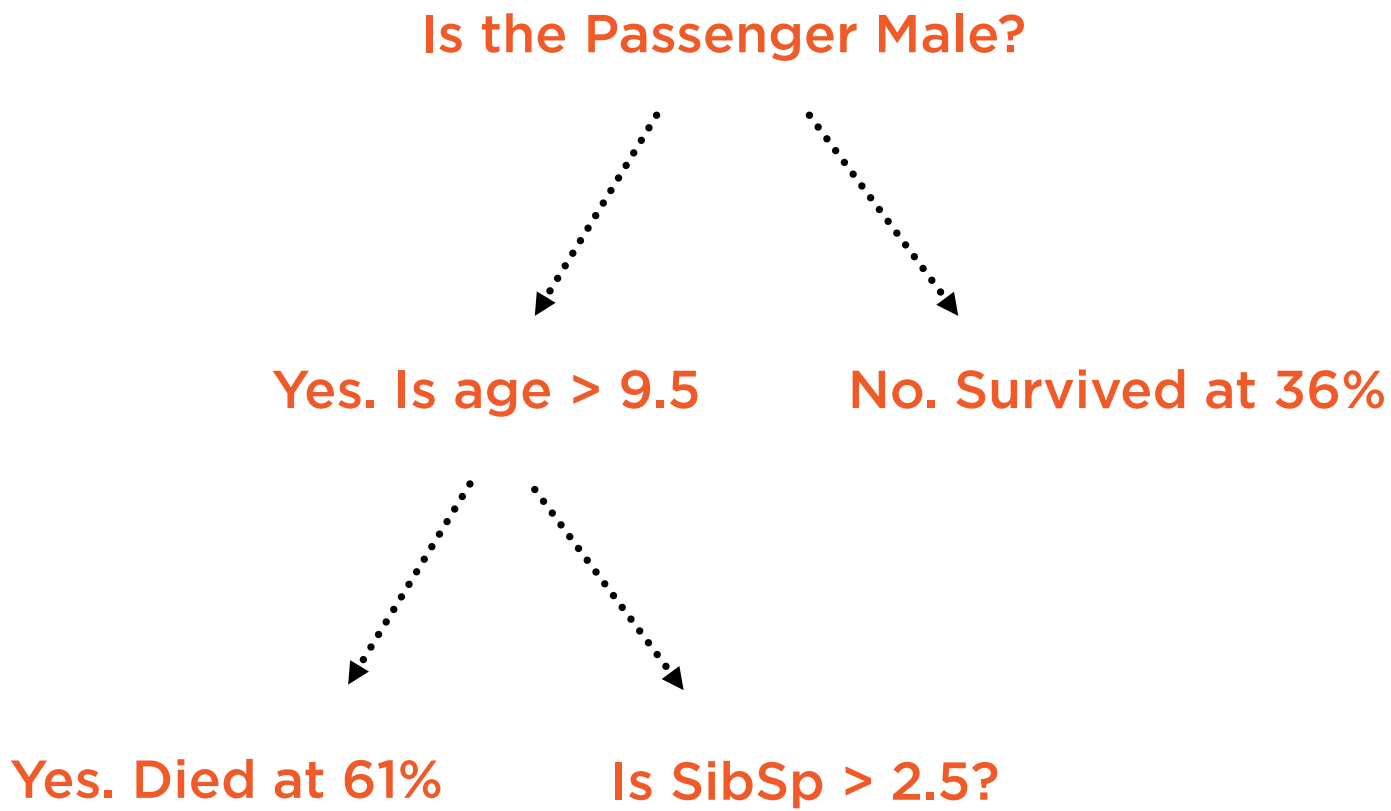
Machine learning basics

Basic math and statistics

Decision Tree



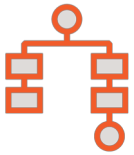
Decision Tree Example



Ensemble Models

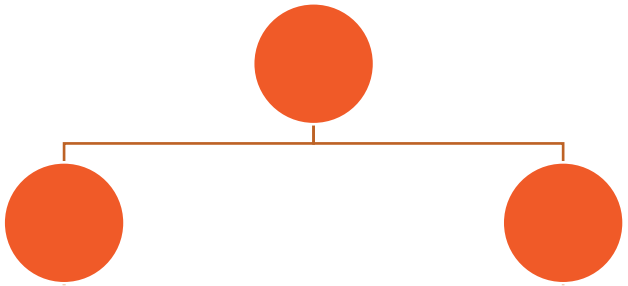
XGBoost has won more Kaggle competitions than any other model in the structured data category

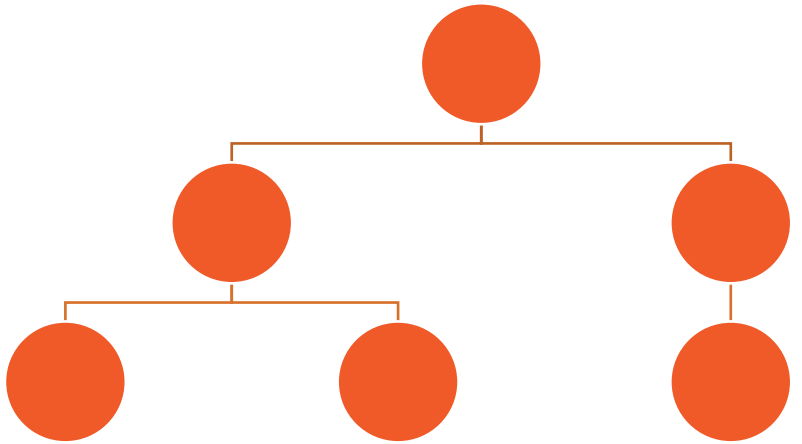
Most ensemble models use the same base learner. In XGBoost that base model is a decision tree



A weak learner is a model that's able to predict something slightly better than 50%

Ensemble Model

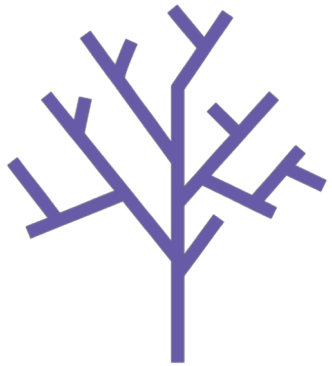




Gradient Boosting

Base Learners

A model that is slightly better than a coin toss



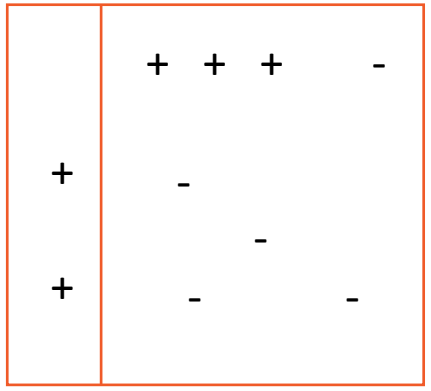
Decision Tree

XGBoost uses decision trees as the base learner

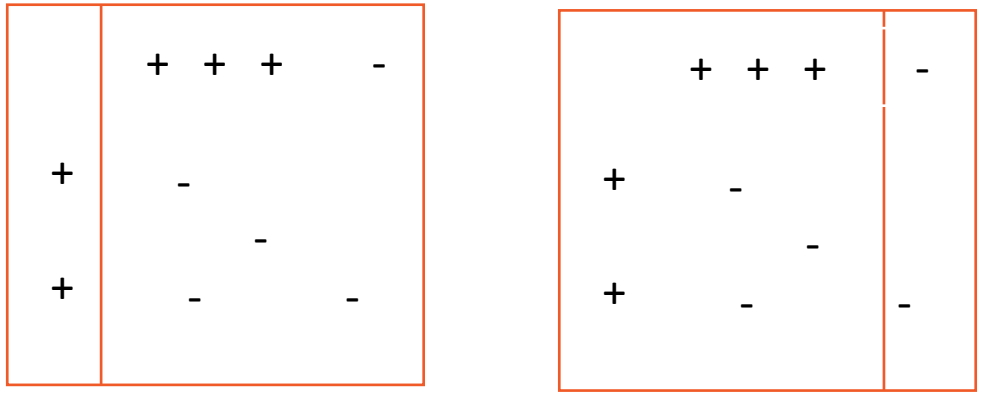
Decision Stumps

A decision stump is a shallow decision tree with little branching

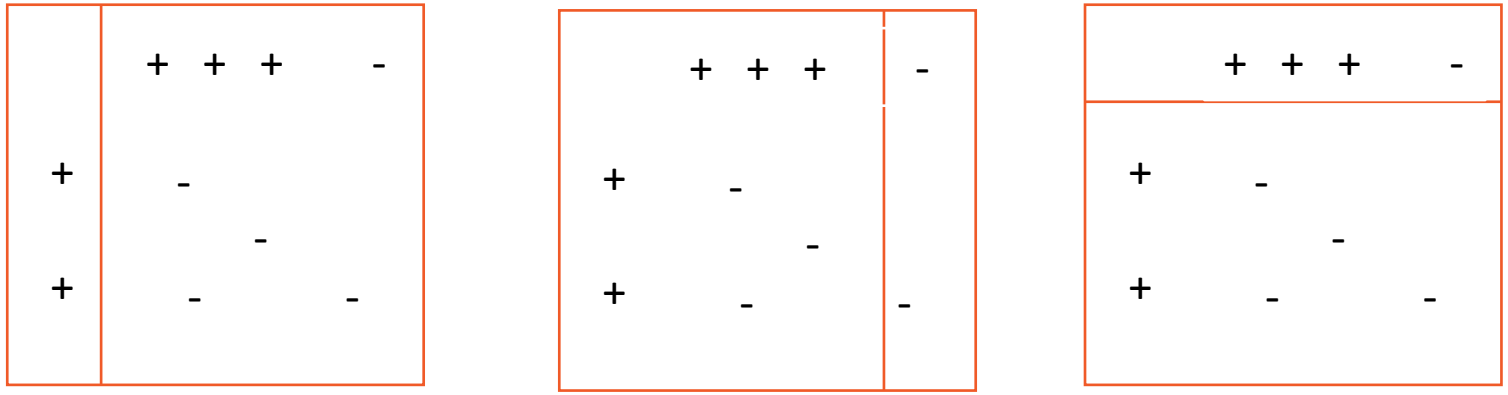
Gradient Boosting



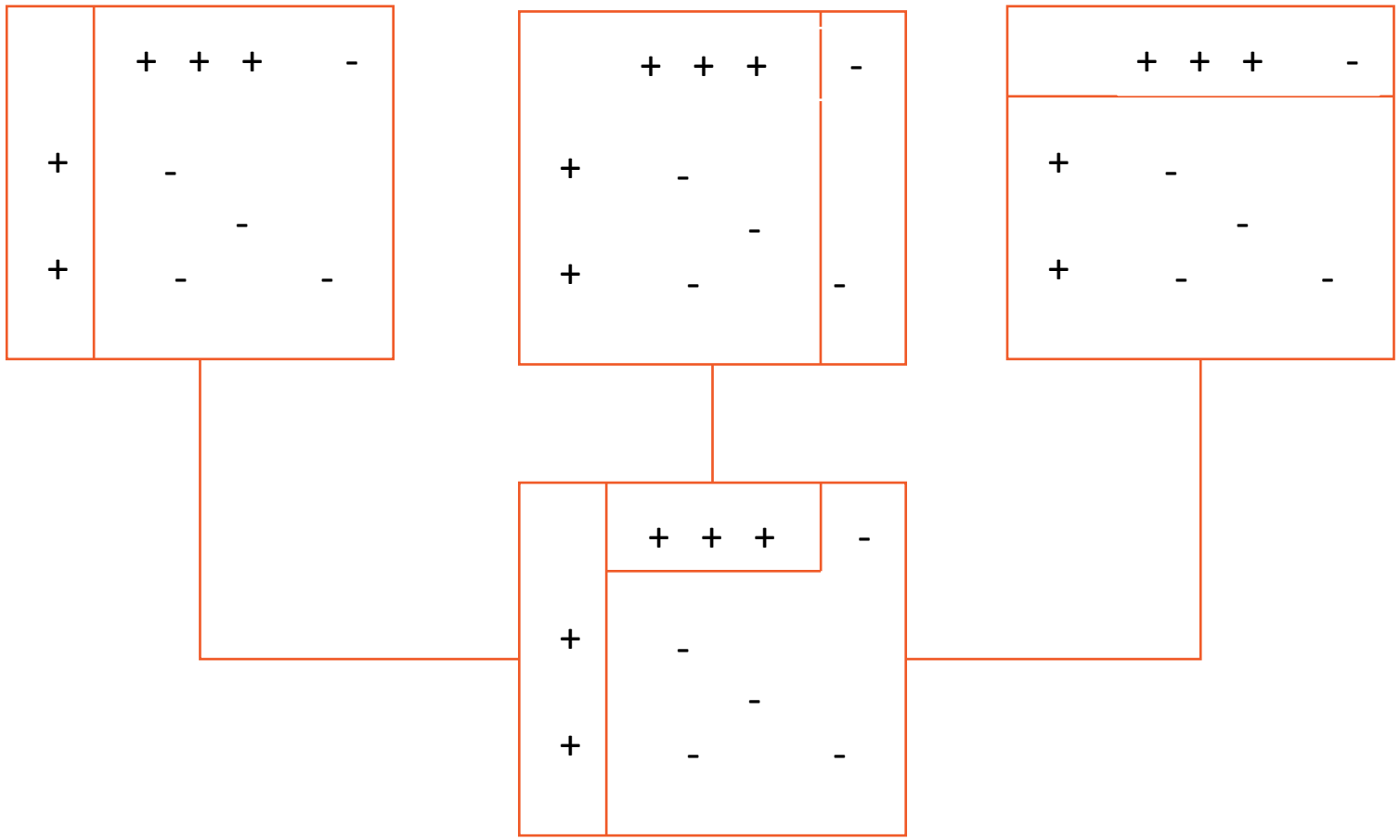
Gradient Boosting



Gradient Boosting



Gradient Boosting



Demo

Import libraries

Load Titanic dataset

Train the XGBoost model

Evaluate model using SciKit-Learn

Score the models accuracy

Summary

- Defined decision tree
- Skills needed for this course
- Ensemble modeling
- Examined gradient boosting
- Completed a demo using XGBoost