Boosting: Gradient Boosting & XGBoost

Hunter Glanz

OUTLINE

Foundational Machine Learning

- You've learned about:
  - Traditional Regression
  - Logistic Regression
  - K-Nearest Neighbors
  - Discriminant Analysis
  - Support Vector Machines
  - Tree-Based Methods

Foundational Machine Learning

- You've learned about:
  - Traditional Regression
  - Logistic Regression
  - K-Nearest Neighbors
  - Discriminant Analysis
  - Support Vector Machines
  - Tree-Based Methods

Remember *there's no free lunch*!

Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!

Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!
  - Near infinite number of ways to do this so we'll talk generally about three broad strategies:
    1. Bagging
    2. Stacking
    3. Boosting

Ensemble Learning Strategies

- Ensemble learning refers to algorithms that combine the predictions from two or more models:
  - Let's team up!
  - Near infinite number of ways to do this so we'll talk generally about three broad strategies:
    1. Bagging
    2. Stacking
    3. Boosting

Today we will focus on **Gradient Boosting** and **XGBoost**

Motivation

A procedure that combines the outputs of many "weak" learners to produce a powerful "committee."

Motivation

A procedure that combines the outputs of many "weak" learners to produce a powerful "committee."

Similar to all boosting methods, Gradient Boosting looks to consecutively reduce error with each consecutive model

Gradient Boosting

- After the initial model is fit, a loss function is plotted (instead of updating the weights as we did in AdaBoost.M1)

Gradient Boosting

- After the initial model is fit, a loss function is plotted (instead of updating the weights as we did in AdaBoost.M1)
- Gradient Boosting gets its name from **Gradient Descent**, which is the method used to find the parameters which minimize the loss function
  - As gradient descent continues along the loss function, it continuously tunes the parameters until the minimum point is found
  - The goal: find the optimal parameters which have the biggest decrease on the loss function
  - By sequentially minimizing our loss function, our model gets stronger and stronger

Small trees are popular choices for the weak learners, despite technically being able to use this with many different methods

XGBoost

e**X**treme **G**radient **Boost**ing

- Direct application of Gradient Boosting for decision trees with the following advantages:
  1. Easy to use
  2. Computational Efficiency
  3. Model Accuracy
  4. Feasibility – easy to tune parameters and modify objectives

The GOAT

The go-to model of kaggle winners