Gradient Boosting Algorithms

Epoch IIT Hyderabad

Himani Agrawal
MA22BTECH11008

Contents

Introduction

Gradient Boosting is a popular machine learning algorithm that belongs to the supervised learning technique. It is a boosting method that combines several weak learners into strong learners. Each new model is trained to minimize the loss function of the previous model using gradient descent.

Working of Algorithm

1. Initialization: The algorithm starts by predicting a constant value for all instances in the dataset. This initial prediction can be the mean (for regression problems) or the most frequent class (for classification problems).
2. Gradient Calculation: The residuals (differences between the predicted and actual values) are calculated. These residuals are the negative gradient of the loss function.
3. Weak Learner Construction: A decision tree is fitted to the residuals from the previous step. This tree is a 'weak learner' and is designed to predict a small step towards the residuals.
4. Prediction Update: The predictions are updated by adding the predictions from the new weak learner, scaled by a learning rate.
5. Iterative Learning: Steps 2-4 are repeated until a specified number of weak learners have been created, or if the residuals

can no longer be reduced.Each iteration involves the following formula

$$F_{i+1} = F_i - f_i$$

Where:

- F_i is the strong model at step i.
- f_i is the weak model at step i.

In each iteration, the algorithm computes the gradient of the loss function with respect to the predictions of the current ensemble and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.

Formulas used

The Gradient Boosting algorithm uses the following formulas:

- Residual Calculation: $r_i = y_i - F(x_i)$

Where r_i is the residual for the ith instance, y_i is the actual value, and $F(x_i)$ is the predicted value.

- Weak Learner Prediction: $h_i(x) = r_i$

Where $h_i(x)$ is the prediction of the ith weak learner.

- Prediction Update: $F_{i+1}(x) = F_i(x) + \eta * h_i(x)$

Where $F_{i+1}(x)$ is the updated prediction, $F_i(x)$ is the current prediction, η is the learning rate, and $h_i(x)$ is the prediction of the ith weak learner