Boosting

Danila Kokin & Haonan Jin & Hang Yao

April 30th 2025

Abstract

Boosting algorithms are a highly effective class of ensemble methods that build powerful predictors by sequentially combining weak learners. This project examines key aspects of boosting, focusing theoretically on the relationship between boosting algorithms, particularly `AdaBoost`, and the concept of maximizing classification margins. This margin-based perspective provides crucial insights into boosting's strong generalization performance and resistance to overfitting. Practically, the project demonstrates the successful application and implementation of boosting for multi-class classification problems using the standard **Iris dataset**, comparing `AdaBoost.M1` and `Gradient Boosting` (GBM) in **R**. Results show high accuracy across these methods for this task. The work highlights the theoretical basis related to margin maximization and confirms the practical utility and ease of implementation of boosting for common multi-class scenarios.

Contents

1 Introduction, Significance, and Purpose

Ensemble methods, techniques that aggregate predictions from multiple machine learning models, often achieve better performance than single models alone. Among these, boosting algorithms have gained importance for their sequential learning process: they iteratively train simple models ("weak learners"), each new learner giving increased focus to examples misclassified by the ensemble built so far. This adaptive re-weighting allows boosting to construct highly accurate final predictors, often reaching state-of-the-art performance on various tasks.

Significance: Understanding why boosting is so effective, especially its ability to avoid overfitting where traditional complexity-based bounds might predict. Otherwise, is crucial for both applying it confidently and for developing improved algorithms. The theory connecting boosting to the maximization of classification margins (a measure of prediction confidence) offers an outstanding explanation. Furthermore, while foundational boosting algorithms focused on binary classification, many real-world problems involve multiple categories, making the effective application of boosting to multi-class scenarios highly significant.

Purpose & Research Questions: This project aims to explore these facets of boosting through targeted theoretical analysis and practical demonstration. Specifically, we focus on:

Theoretically analyzing the relationship between boosting algorithms (specifically `AdaBoost's` mechanism) and the concept of maximum margin classifiers, exploring how boosting implicitly works towards maximizing prediction confidence.

Practically implementing and evaluating boosting algorithms (`AdaBoost.M1`, `GBM`) designed to handle multi-class classification directly, using the standard **Iris dataset** as a case study. Through this focused investigation, we seek to clarify the margin perspective on boosting and demonstrate its practical effectiveness in a common multi-class problem.

2 Literature Review

The development of boosting algorithms began with `AdaBoost`, introduced by **Freund** and **Schapire**, which provided both a practical algorithm and initial theoretical guarantees based on the "weak learning" hypothesis. `AdaBoost's` empirical success, especially its often-observed resistance to overfitting, prompted deeper theoretical investigation. While early bounds based on the number of combined classifiers predicted overfitting, the "margins explanation" emerged as a more fitting theory. **Schapire**, **Freund**, **Bartlett**, and **Lee** showed that `AdaBoost` tends to increase the margins of training examples, even after zero training error is reached. Generalization bounds based on these margins (like

Chapter 5 in Schapire and Freund [2012]) demonstrated that error on unseen data depends more on the margin distribution and base learner complexity than on the number of boosting rounds, thus explaining the lack of overfitting. **Friedman** later generalized boosting with `Gradient Boosting Machines` (GBM), framing it as functional gradient descent, allowing for various loss functions and improving robustness over `AdaBoost's` exponential loss Fernández [2025]. For multi-class problems, extensions of `AdaBoost` and native implementations within `GBM` to handle more than two categories effectively.

3 Theoretical Questions

This section theoretically addresses the relationship between boosting and maximum margin classifiers. The discussion is grounded in established concepts from optimization (specifically, minimizing exponential loss) and statistical learning theory (margin definitions), referencing the foundational work.

3.1 Boosting and Maximum Margin Classifiers

To understand the relationship between boosting and maximum margin classifiers, we begin with the objective of boosting, which is to minimize the **exponential loss** in binary classification ($y \in \{-1, +1\}$):

$$L(y, f(x)) = \exp(-yf(x)), \tag{1}$$

where $f(x)$ is the score predicted by the classifier. Boosting builds $f(x)$ as a linear combination of weak learners:

$$f(x) = \sum_{m=1}^{M} \beta_m G_m(x), \tag{2}$$

where β_m is the coefficient and $G_m(x)$ is the prediction of the m-th weak learner. The goal is to minimize the total loss over the training set:

$$J(f) = \sum_{i=1}^{N} \exp(-y_i f(x_i)) \tag{3}$$

In this context, the **margin** of a training example (x_i, y_i) under the combined classifier (Equation 2) is defined as:

$$\text{margin}(x_i, y_i) = \frac{y_i f(x_i)}{\sum_m |\beta_m|} = \frac{y_i \sum_m \beta_m G_m(x_i)}{\sum_m |\beta_m|}. \tag{4}$$

The margin quantifies the confidence of the classification: a large positive margin implies a confident, correct prediction, while a negative margin indicates misclassification.

Although `AdaBoost` minimizes the exponential loss (Equation 1) rather than directly optimizing the margin, it implicitly maximizes margins. The exponential loss should penalize negative margins more heavily than positive ones, since positive margin observations are already correctly classified.

As shown by **Schapire et al.**, `AdaBoost's` iterative reweighting mechanism prioritizes misclassified or low-margin examples, pushing the decision boundary to achieve larger margins. This behavior helps explain `AdaBoost's` strong generalization performance, even when training error is reduced to zero (Hastie et al. [2009], Schapire and Freund [2012]).

4 Practical Implementation Strategy

We outline algorithms and experimental designs using common libraries (**R's** libraries). This involves describing algorithm, data simulation for experiments, visualization code structure (*ggplot2* in **R**), performance metric calculation.

4.1 Multiclass Boosting

Standard `AdaBoost` was originally designed for binary classifications. This task asks us to show how boosting can handle problems with more than two categories, for example, in our case we use the **Iris dataset** in order to classifying an iris flower as *Setosa*, *Versicolor*, or *Virginica*. We wanna compare what are the results with the different algorithms of boosting that can handle multi-classification problems.

R Code *multiclass_boosting_example.R* :

- **Data :** We used the standard **Iris dataset**, which has 3 classes (Species) and 4 features (Sepal.Lenght, Sepal.Width, Petal.Lenght,Petal.Width).
- **Models & Implementations:** We have implemented 2 algorithms to try.
  - **AdaBoost.M1:** Is the foundational boosting algorithm. It works by training a sequence of weak learners. `AdaBoost.M1` is an early adaptation of `AdaBoost` designed to handle multi-class problems (ada [n.d.]).
  - **GBM:** `GBM` is a generalization of boosting framed as an optimization problem using gradient descent in function space. What it differs from `Adaboost` is that

this one focuses on minimizing a loss function through a gradient descent-like process, using the residuals or pseudo-residuals of previous iterations as training data for the next weak learner (Fernández [2025] , Wikipedia contributors [2025]).

- **Evaluate:** We predicted the classes on the test set for each model and computed accuracy using a confusion matrix.

In Figure 1, we observe that both models achieved the same accuracy of 93.3%, indicating comparable performance in classifying the dataset correctly. Despite their methodological differences, the identical accuracy suggests that neither model significantly outperformed the other in terms of overall predictive capability.

This similarity in performance may be attributed to the simplicity of the dataset. With only 150 observations, 4 input variables, and just 3 target classes (species), the classification task is relatively straightforward, making it easier for multiple models to achieve high accuracy.

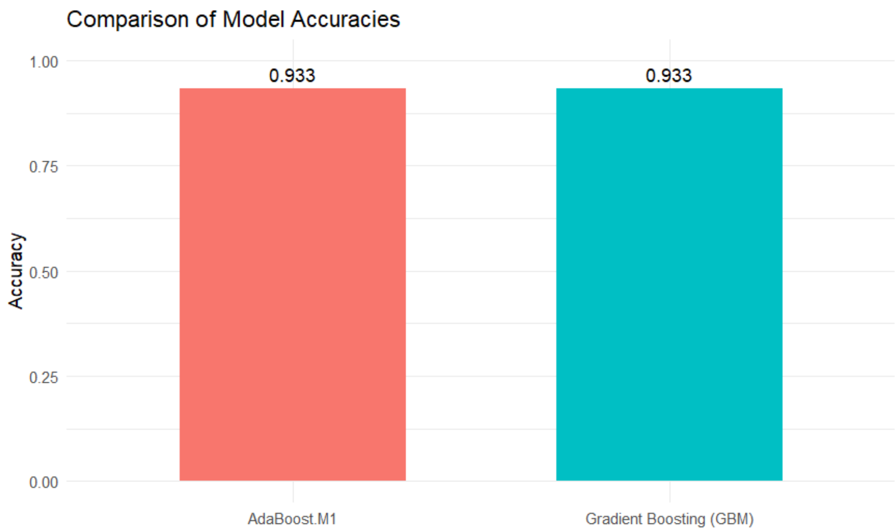

Figure 1: Boosting Multiclass

5 Interpretation and Discussion

This project explored boosting through a specific theoretical and a practical application. Theoretically, we examined the connection between boosting and margin maximization. This perspective is crucial because it helps explain one of boosting's most notable properties, its resistance to overfitting, even when many weak learners are combined. Unlike complexity measures that grow with the number of boosting rounds, margin-based bounds (discussed conceptually in the literature review) depend on the confidence of predictions on the training data. `AdaBoost's` mechanism, by focusing on misclassified or low-margin points via its weight updates (implicitly linked to minimizing exponential loss), acts to increase these margins, thereby often improving generalization even late into the training process.

Practically, we investigated the application of boosting to multi-class classification using the **Iris dataset**. The results showed that different boosting algorithms (`AdaBoost.M1`, `GBM`) achieved identical high accuracy (93.3%) on this specific task. This suggests that for simpler, well-behaved datasets like **Iris**, multiple boosting approaches can easily find near-optimal solutions.

Overall, the theoretical insight into margin maximization provides a strong rationale for boosting's effectiveness, while the practical multi-class implementation confirms its utility and accessibility for standard classification tasks.

6 Formulated Conclusions

Based on the theoretical analysis and practical implementation conducted, we draw the following conclusions:

Boosting algorithms, exemplified by `AdaBoost`, demonstrate a fundamental connection to the principle of margin maximization in classification. By minimizing loss functions like the exponential loss, they implicitly drive up the margins of training examples, offering a theoretical explanation for their observed strong generalization performance and resistance to overfitting. Moreover, boosting is practically effective applicable to multi-class classification problems. Standard libraries in **R** (*adabag*, *gbm*) provide direct mechanisms to handle multiple classes, achieving high accuracy on datasets like **Iris**.

7 Recommendations for Future Work

Theoretical Deep Dive: Fully derive and analyze the margin-based generalization bounds (Chapter 5 from Schapire & Freund in Schapire and Freund [2012]) to gain a quantitative understanding of the trade-offs between margin distribution, base learner complexity, and sample size.

Multi-Class Robustness: Extend the practical multi-class evaluation to larger, more complex datasets and incorporate an analysis of robustness to label noise, comparing the sensitivity of `AdaBoost.M1` and `GBM`, and also include a additional `XGBoost` model in the multi-class setting.

References

Robert E. Schapire and Yoav Freund. *Boosting: Foundations and Algorithms*. The MIT Press, 2012. ISBN 978-0-262-30118-3. doi: 10.7551/mitpress/8291.001.0001. URL https://doi.org/10.7551/mitpress/8291.001.0001. Accessed: 2025-04-21.

Germán Fernández. Boosting: una intuición visual. https://gefero.github.io/flacso_ml/clase_3/notebook/boosting_intuicion_notebook.nb.html, 2025. Accessed: 2025-04-24.

T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning: Data mining, inference, and prediction. Springer (2nd ed.), 2009. URL https://www.sas.upenn.edu/~fdiebold/NoHesitations/BookAdvanced.pdf. Accessed: 2025-04-21.

adabag-package function - rdocumentation. https://www.rdocumentation.org/packages/adabag/versions/5.0/topics/adabag-package, n.d. Accessed: 2025-04-21.

Wikipedia contributors. Gradient boosting — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Gradient_boosting, 2025. Accessed: 2025-04-24.