Advanced Ensemble Techniques

Your Name

2024-07-22

Introduction

In this document, we explore advanced ensemble techniques such as XGBoost, LightGBM, and CatBoost. We will review their theoretical foundations, implement models, and evaluate their performance on real-world datasets.

Theory Review

Summarize the theoretical foundations. For instance, the boosting process can be expressed mathematically as follows:

$$F(x) = \sum_{m=1}^{M} \alpha_m h_m(x)$$

where $F(x)$ is the final model, α_m is the weight of the m-th model, and $h_m(x)$ is the m-th weak learner.

Implementation

- **XGBoost**: Implementation steps for XGBoost.
- **LightGBM**: Implementation steps for LightGBM.
- **CatBoost**: Implementation steps for CatBoost.

Experimentation

Describe the datasets and benchmarking. For instance, compare model performance using metrics such as accuracy (Acc), precision (Prec), and recall (Rec).

$$\text{F1 Score} = 2 \times \frac{\text{Prec} \times \text{Rec}}{\text{Prec} + \text{Rec}}$$

Analysis

Evaluate efficiency and scalability. Discuss findings in terms of training time, model accuracy, and resource utilization.

Conclusion

Summarize key findings, comparing the performance of XGBoost, LightGBM, and CatBoost, and provide any conclusions.