Light Gradient Boosting Machine

Aashish Mahato

Agenda

+ Gradient Boosting
+ Light Gradient Boosting Machine

Gradient Boosting

What is Gradient Boosting?

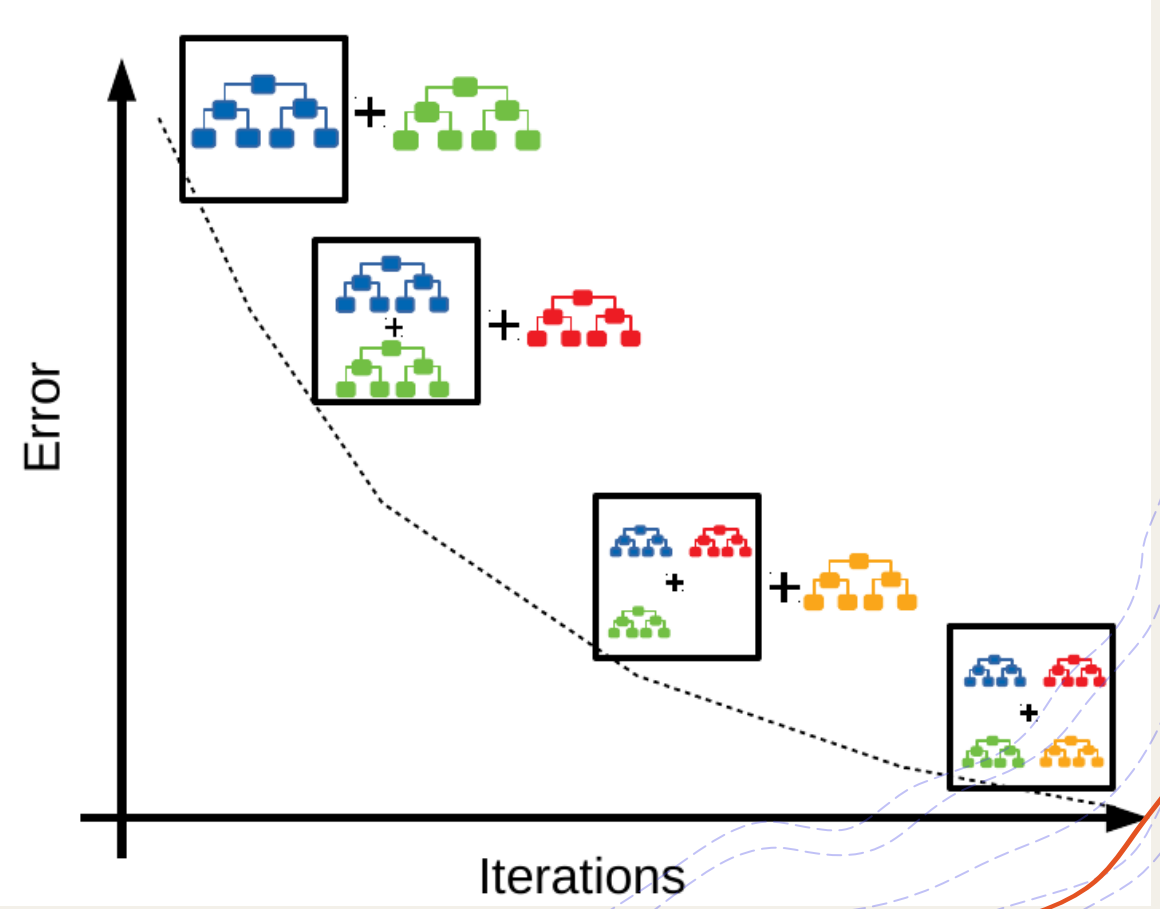
+ Gradient boosting is an ensemble ML technique that combines a collection of weak learners into a single, more accurate and efficient predictive model.

+ A weak learner is a machine learning model that is slightly better than a random guessing model. For example, let's say we are classifying mushrooms into edible and inedible. If a random guessing model is 40% accurate, a weak learner would be just above that: 50-60%.

+ These weak learners are typically decision trees, which is why the algorithms are commonly referred to as gradient boosted decision trees (GBDTs).

+ Gradient boosting algorithms work iteratively by adding new models sequentially, with each new addition aiming to resolve the errors made by the previous ones.

+ The final prediction of the aggregate represents the sum of the individual predictions of all the models. Gradient boosting combines the gradient descent algorithm and boosting method, with a nod to each component included in its name.



Light Gradient Boosting Machine (LightGBM)

What is LightGBM?

+ LightGBM is a powerful and efficient open-source gradient boosting framework developed by Microsoft. It's specifically designed to handle large datasets and perform well in terms of speed and memory usage.

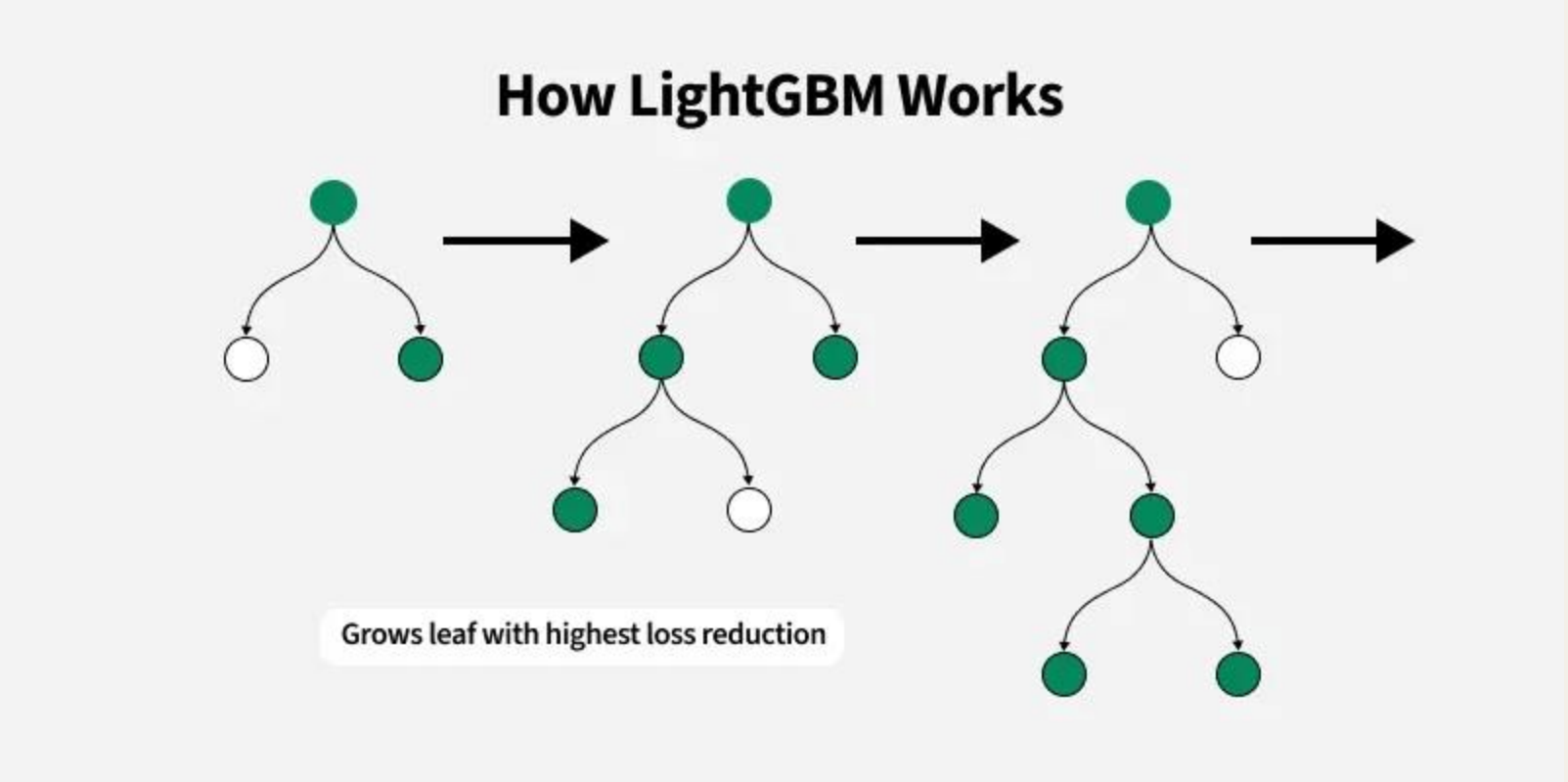
+ LightGBM employs a leaf-wise decision tree-based gradient boosting method that reduces memory usage while enhancing model efficiency. This method adopts two innovative techniques, Gradient-based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB).

Gradient-based One-Side Sampling (GOSS)

+ Instead of using all the data instances to find the optimal split points for a tree, GOSS prioritizes instances with larger gradients (i.e., those that have a higher error and are more difficult to predict). It randomly samples instances with smaller gradients. This approach helps the model focus on the more challenging examples, leading to faster training without a significant loss in accuracy.

Exclusive Feature Bundling (EFB)

+ This technique reduces the number of features by grouping together mutually exclusive features that are rarely, if ever, active at the same time. This can significantly speed up training by reducing the computational cost of finding the best split for each tree.



- LightGBM grows tree vertically while other tree based learning algorithms grow trees horizontally.
- It means that LightGBM grows tree leaf-wise while other algorithms grow level-wise.
- It will choose the leaf with max delta loss to grow. When growing the same leaf, leaf-wise algorithm can reduce more loss than a level-wise algorithm.

Advantages of LightBGM

+ Faster training speed and higher accuracy: It outperforms other gradient boosting algorithms on training speed and accuracy.

+ Low memory usage: Optimized for memory efficiency and handling large datasets with minimal overhead.

+ Parallel and GPU learning support: Takes advantage of multiple cores or GPUs for faster training.

+ Effective on large datasets: Its optimized techniques such as leaf-wise growth and histogram-based learning make it suitable for big data applications.

Disadvantages of LightBGM

+ Overfitting: Light GBM split the tree leaf-wise which can lead to overfitting on small datasets as it produces much complex trees.

Feature Importance

What is Feature Importance?

+ Feature importance refers to techniques that calculate a score for all the input features for a given model. The scores represent the “importance” of each feature.

+ The scores are useful and can be used in a range of situations in a predictive modeling problem, such as:

  - Better understanding the data.
  - Better understanding a model.
  - Reducing the number of input features.

SHAP (SHapley Additive exPlanations)

What are SHAP Values?

+ SHAP (SHapley Additive exPlanations) values are a way to explain the output of any machine learning model. It uses a game theoretic approach that measures each player's contribution to the final outcome. In machine learning, each feature is assigned an importance value representing its contribution to the model's output.

+ SHAP values show how each feature affects each final prediction, the significance of each feature compared to others, and the model's reliance on the interaction between features.

+ Features with positive SHAP values positively impact the prediction, while those with negative values have a negative impact. The magnitude is a measure of how strong the effect is.

Thank You