LightGBM Algorithm

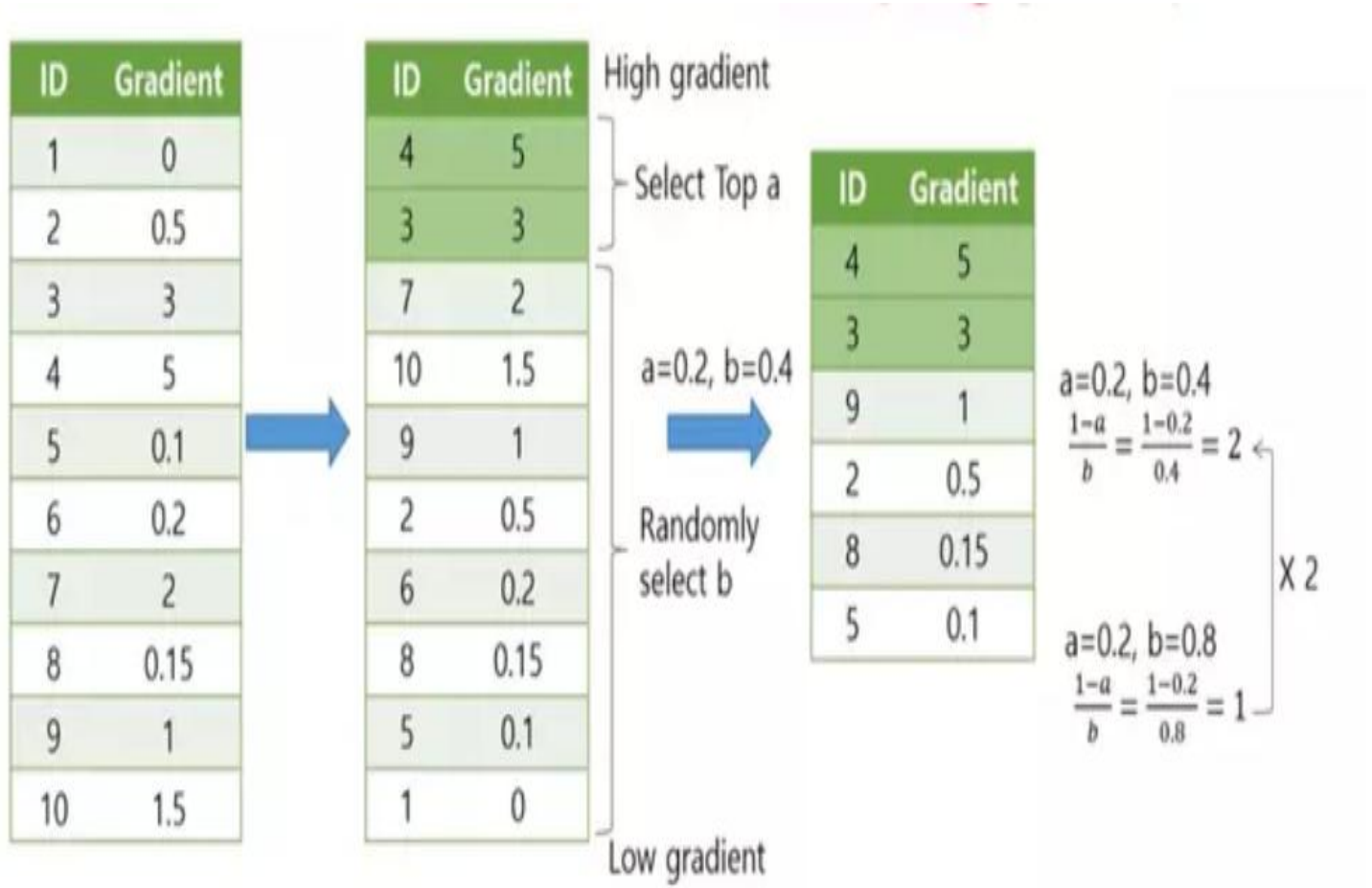
Light Gradient Boosting Machine-LG BOOST

LightGBM is most featured by its faster training efficiency and great accuracy.

Special techniques are:

- Gradient-based One-Side Sampling(GOSS)
- Histogram Based Best Value Search in Tree Node Splitting
- Exclusive Feature Bundling
- Leaf-wise Tree Growth Strategy

1.Gradient-based One-Side Sampling(GOSS)

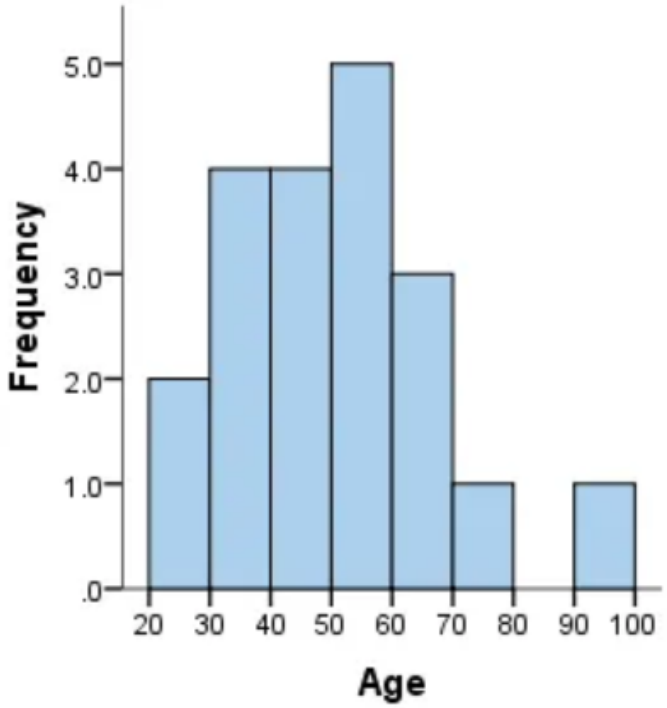


2. Histogram Based Best Value Search in Tree Node Splitting

LightGBM splitting criteria is to reduce gradient variance from parent to children.

For example, given the below age feature, histogram discrete feature values into different range bins, so we can use spiting criteria like Age⩽30, Age⩽40, , , Age⩽100 instead of trying all the possible age value like Age⩽31, Age⩽32 etc.

Parent Node Histogram = Left Child Histogram + Right Child Histogram



3.Exclusive Feature Bundling

EFB aims to reduce features by merging features, specifically, merge the mutually exclusive features, which rarely take non-zero values at the same time.

LightGBM provides the below two algorithms to implement

- Identify mutually exclusive feature bundles from training set
- Merge feature bundle and assign a value to the bundle

Steps for EFB

Step 1: Original Sparse Dataset

Consider a dataset with 4 sparse features:

Sample	Feature 1	Feature 2	Feature 3	Feature 4
1	1	0	0	0
2	0	1	0	0
3	0	0	1	0
4	0	0	0	1
5	1	0	0	0

Step 2: Identify Non-Overlapping Features

Using EFB, we group features that rarely or never have non-zero values at the same time.
In this example:
- Feature 1 and Feature 2 do not overlap
- Feature 3 and Feature 4 do not overlap

Step 3: Create Feature Bundles

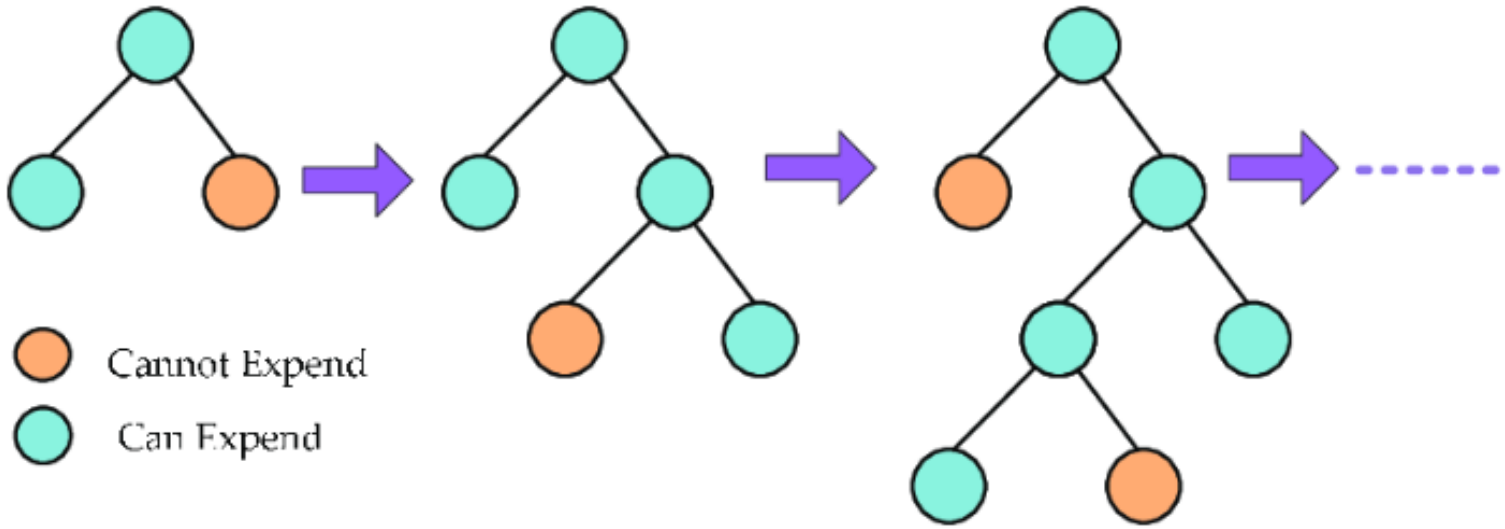
The non-overlapping features can be bundled into two groups:
- Bundle 1: Feature 1 and Feature 2
- Bundle 2: Feature 3 and Feature 4

This reduces the total number of features from 4 to 2.

The bundled features are represented as follows:

Sample	Bundle 1 (Feature 1 & 2)	Bundle 2 (Feature 3 & 4)
1	1 ($from\ Feature\ 1$)	0
2	2 ($from\ Feature\ 2$)	0
3	0	1 ($from\ Feature\ 3$)
4	0	2 ($from\ Feature\ 4$)
5	1 ($from\ Feature\ 1$)	0

4.Leaf-wise Tree Growth Strategy



Leaf-wise tree expansion in LightGBM

Difference b/w XG and LG BOOST