LightGBM

Mannem Charan AI21BTECH11019

Abstract

This report consists of my basic understanding of one of the popular Ml methods "Light-GBM".

1 LIGHTGBM

LightGBM,Light Gradient Boosting Machine is a boosting algorithm which is a modification of it's ancestor GBM.It combines the decision trees in an effective manner which improves the performance of model and reduces the memory usage.It uses two novel techniques,

1) Gradient-based One Side Sampling (GOSS)
2) Exclusive Feature Bundling (EFB)

They comprise together to make the model work efficiently and provide it a cutting edge over other GBDT frameworks.

2 GRADIENT-BASED ONE SIDE SAMPLING

Different data instances have varied roles in the computation of information gain. The instances with larger gradients(i.e., under-trained instances) will contribute more to the information gain. GOSS keeps those instances with large gradients and only randomly drop those instances with small gradients to retain the accuracy of information gain estimation. This treatment can lead to a more accurate gain estimation than uniformly random sampling, with the same target sampling rate, especially when the value of information gain has a large range.

3 EXCLUSIVE FEATURE BUNDLING

High-dimensional data are usually very sparse which provides us a possibility of designing a nearly lossless approach to reduce the number of features. Specifically, in a sparse feature space, many features are mutually exclusive, i.e., they never take nonzero values simultaneously. The exclusive features can be safely bundled into a single feature (called an Exclusive Feature Bundle). Hence, the complexity of histogram building changes from O(#data × #feature) to O(#data × #bundle), while #bundle << #feature . Hence, the speed for training framework is improved without hurting accuracy.

4 ADVANTAGES OF LightGBM

1) Faster training speed and higher efficiency
2) Lower memory usage
3) Better accuracy than any other algorithm
4) Capability with Large Datasets.