Report

Tree boosting is a highly effective and widely used machine learning method. scalable end-to-end tree boosting system called XGBoost, which is used widely by data scientists to achieve state-of-the-art results on many machine learning challenges. Machine learning and data-driven approaches are becoming very important In many areas. Smart spam classifiers protect our email by learning from massive amounts of spam data and user feedback; advertising systems learn to match the right ads with the right context; fraud detection systems protect banks from malicious attackers; anomaly event detection systems help experimental physicists to find events that lead to new physics.

Advantages: Advantages of XGBoost Algorithm in Machine Learning. XGBoost is an efficient and easy to use algorithm which delivers high performance and accuracy as compared to other algorithms. XGBoost is also known as a regularized version of GBM.

Disadvantages: XGBoost does not perform so well on sparse and unstructured data. A common thing often forgotten is that Gradient Boosting is very sensitive to outliers since every classifier is forced to fix the errors in the predecessor learners. The overall method is hardly scalable.

Applications: A wide range of applications: Can be used to solve regression, classification, ranking, and user-defined prediction problems. 1)Portability: Runs smoothly on Windows, Linux, and OS X. 2)Languages: Supports all major programming languages including C++, Python, R, Java, Scala, and Julia. 3)Cloud Integration: Supports AWS, Azure, and Yarn clusters and works well with Flink, Spark, and other ecosystems

Conclusion: Building XGBoost, a scalable tree boosting system that is widely used by data scientists and provides state-of-the-art results on many problems. We proposed a novel sparsity aware algorithm for handling sparse data and a theoretically justified weighted quantile sketch for approximate learning.