Wyatt Blair

EMT-678-WS: Big Data Technologies

11/1/2024

**Class XGBoost: A Scalable Tree Boosting System**

XGBoost is a boosting technique which allows for powerful and scalable performance enhancement on classification and regression problems. The technique uses an ensemble of weak learners which are trained sequentially. Each model improves on the error of its predecessor by paying extra mind to the data-points which caused its predecessor problems. By employing defining a “block system” to apply to large datasets some run-time advantages can be made. When performing a sorting algorithm (which is essential when training trees so an appropriate split can be made), it is more time-efficient to approach each “block” of data individually rather than trying to sort the entire dataset. The authors point to several competitions and datasets where XGBoost proves to be the best tool for the job.