

Investigating Optimization Techniques for GLM in Various Frameworks

Generalized Linear Models (GLMs) are commonly used in statistical modelling, with different frameworks employing unique optimization methods to estimate parameters effectively. The table below provides a comparative overview of these techniques across various platforms.

Module/Framework/Package	Algorithm Description	Example of Superior Performance
Base R (stats package)	Iteratively Reweighted Least Squares (IRLS) is used for GLM estimation.	Suitable for small to moderately sized datasets that fit into memory.
Big Data R (High Performance Computing)	Packages like bigglm handle large datasets by processing data in chunks, reducing memory usage.	Effective when dealing with datasets too large to fit into memory, providing better performance than base R.
Dask-ML	Implements scalable algorithms compatible with Dask's parallel computing, allowing distributed processing of large datasets.	Outperforms scikit-learn when handling large datasets that benefit from parallel and distributed computing.
SparkR	Utilizes Spark's distributed computing capabilities to fit GLMs on large datasets across clusters.	Excels in scenarios requiring distributed data processing, offering superior performance over base R for large-scale data.
Spark Optimization (MLlib)	Employs optimization techniques like Gradient Descent and L-BFGS for scalable machine learning, including GLMs.	Ideal for large-scale data processing, providing efficient model training compared to single-machine implementations.
scikit-learn	Uses methods such as Ordinary Least Squares for linear regression and logistic regression for binary classification.	Suitable for datasets that fit into memory; for larger datasets, frameworks like Dask-ML or Spark MLlib are more appropriate.

IRLS remains a staple for traditional GLMs, but parallelized or distributed approaches like Spark's LBFGS are crucial for large datasets. Gradient-based techniques (SGD, LBFGS) excel in processing massive data efficiently, ensuring scalability. Meanwhile, regularization methods such as Elastic Net and Coordinate Descent are particularly effective for high-dimensional data, where feature selection plays a key role.