Exploring Optimization Methods in GLM Across Different Frameworks

Generalized Linear Models (GLMs) are widely used in statistical modeling, and different frameworks implement distinct optimization techniques to estimate parameters efficiently. Below is a comparative summary of the optimization techniques across various frameworks.

Framework/Package	Optimization Method	Brief Description	Use Case
Base R (stats package)	Iteratively Reweighted Least Squares (IRLS)	A weighted least squares approach that updates weights iteratively to estimate GLM parameters.	Works well for small to moderate datasets where traditional GLMs apply without regularization.
Big Data R (High-Performance Computing R)	Parallelized IRLS	Similar to IRLS but optimized for distributed computing.	Best suited for large-scale GLM problems in high-performance computing environments.
Dask ML	Stochastic Gradient Descent (SGD)	Iteratively updates model parameters using gradients to minimize the error.	Ideal for handling large datasets that do not fit in memory.
Spark R (MLlib in Apache Spark)	LBFGS (Limited-memory BFGS)	An optimization algorithm for convex functions, useful for high-dimensional data.	Preferred for big data applications requiring distributed computation.
Spark Optimization (MLlib)	Gradient Descent Variants	Implements different forms of gradient descent for optimizing GLM parameters.	Effective in large-scale machine learning workflows across multiple nodes.
Scikit-learn	Coordinate Descent & Elastic Net	A regularization-based method that updates coefficients iteratively to improve prediction accuracy.	Best suited for high-dimensional sparse datasets, such as text classification or genomic data.

In brief IRLS remains a classic choice for traditional GLMs, but when handling large datasets, parallelized or distributed methods (e.g., Spark's LBFGS) become essential. Gradient-based methods (SGD, LBFGS) are powerful when dealing with vast amounts of data that require scalable solutions. Regularization methods (Elastic Net, Coordinate Descent) shine in high-dimensional data where feature selection is crucial.