

GLM Optimization Method Comparison Table

Module/Framework/Package	Name & Description of the Algorithm	Example Where It's Superior to Base R (Identify Python Equivalent)
Base R (glm in stats)	Uses Iteratively Reweighted Least Squares to estimate parameters via maximum likelihood. Efficient for small to medium datasets but lacks parallelization.	The model faces challenges when dealing with very large datasets because of memory constraints. The equivalent Python implementation uses statsmodels.api.GLM which implements Iteratively Reweighted Least Squares.
Big Data R (bigglm in biglm package)	A streaming-based IRLS approach is implemented to handle large datasets that do not fit in RAM. It uses incremental updates rather than loading all data at once.	The best tool for handling millions of patient records in healthcare when a dataset is too large for Base R. Equivalent in Python: Scikit-learn's SGDClassifier (which also processes data incrementally).
Dask ML (dask-ml.linear_model.LinearRegression & dask-ml.glm package)	The algorithm implements Stochastic Gradient Descent (SGD) and Newton's Method which are suitable for distributed computing. The system allows out-of-core learning when the dataset exceeds memory capacity.	Works better when training a large logistic regression model on 100+ million transactions in financial fraud detection. Equivalent in Python: Scikit-learn's SGDClassifier, but Dask processes data in parallel.
Spark R (spark.glm in SparkR package)	Uses L-BFGS (Limited-memory BFGS), which is an iterative, memory-efficient optimization method. Optimized for	Outperforms Base R for real-time analytics on streaming healthcare claims data, where processing speed is crucial.

	distributed Spark environments.	Equivalent in Python: Scikit-learn's SGDClassifier running on a Spark cluster.
Spark MLlib Optimization (mllib.optimization in Spark MLlib)	Implements L-BFGS and SGD in a distributed fashion, making it suitable for massive datasets. L-BFGS is preferred for logistic regression due to its efficiency in high-dimensional problems.	Ideal for predictive modeling on terabytes of e-commerce data, where Base R would crash. Equivalent in Python: Scikit-learn's SGDClassifier with L-BFGS solver.
Scikit-learn (sklearn.linear_model.LogisticRegression and SGDClassifier)	Supports multiple solvers, including Newton's method, L-BFGS, and SGD, providing flexibility for different dataset sizes and computational needs.	More efficient than Base R for handling high-dimensional genetic data (e.g., predicting disease risk from genome sequences). Equivalent in R: glm but with IRLS (less scalable).