HDS 5230: Week 10 Application Assignment's Instructions

For this week's assignment, investigate the details of the optimization method used to find the parameter values of a generalized linear model in each of the following frameworks/packages/modules:

a) Base R (in the stats library)
b) Big data version of R (look here: https://cran.r-project.org/web/views/HighPerformanceComputing.html and here: https://journal.r-project.org/archive/2011-2/RJournal_2011-2_Marschner.pdf)
c) Dask ML: https://dask-glm.readthedocs.io/en/latest/api.html#api-algorithms
d) Spark R: https://spark.apache.org/docs/latest/api/R/spark.glm.html
e) Spark optimization: https://spark.apache.org/docs/latest/ml-advanced.html
f) Scikit-learn: https://scikit-learn.org/stable/modules/linear_model.html#generalized-linear-regression

After going through the documentation, please summarize your findings via a table whose structure matches the one shown below:

Module/framework/package	Name and a brief description of the algorithm	An example of a situation where using the provided GLM implementation provides superior performance compared to that of base R or its equivalent in Python (identify the equivalent in Python)