Project 3: Logistic Regression with ADMM

One large dataset has been partitioned into 10 data files, which are stored here: /gpfs/projects/AMS598/projects2025_data/project3_data/. Each file contains one column called 'y', which is a binary 0/1 response variable, and 25 other columns named x1 – x25, representing explanatory variables.

Requirements:

1. Use the ADMM algorithm to run a logistic regression on all data, and **obtain one set of consensus estimates for all coefficients (intercept and explanatory variables)**.
2. Use mpi4py on SeaWulf for the computation.
3. You can use 10 processes in this project (one process per file), but your program should be able to run with arbitrary number of files.
4. There should be no intermediate files.

Submission:

1. Write a report about the analysis and submit a PDF file about the analysis to BrightSpace.
2. Submit one script file and one slurm file to BrightSpace.
3. Upload the same script file and the same slurm file to the following directory so your analysis and results can be repeated:
   /gpfs/projects/AMS598/projects2025_submission/project3_submission/
   LastName_Firstname /