Project 1

The aim of the project is to implement Iterative Reweighted Least Squares optimization algorithm for logistic regression and perform experiments proving that algorithm works properly.

Part 1

- Implement Iterative Reweighted Least Squares optimization algorithm for parameter estimation in logistic regression.
- Include additional functionality based on the possibility of adding interactions between two variables. The interaction between two variables in logistic regression is understood as the product of these two variables. Add an additional argument: a matrix with two columns, in which each row specifies a pair of variables between which we want to consider interactions.

Part 2 (simulation experiments)

- Develop the methodology of testing the correctness of implementation (create and use artificial data).
- Perform empirical experiments proving that the implemented algorithm works properly.

Part 3 (real datasets experiments)

- Convergence analysis: check how the value of log-likelihood function depends on the number of iterations for 4 above algorithms.
- Find 2 different datasets corresponding to classification problem with binary class variable. You can use UCI repository: https://archive.ics.uci.edu/ or other sources.
- Prepare these datasets to run classification algorithms. This point includes:
    - dealing with missing values,
    - converting categorical variables into numerical ones (use e.g. one-hot encoding),
    - removing collinear variables
    - split data into training and testing sets
- Compare the classification performance of implemented logistic regression with available implementations of this model and 3 popular classification methods: LDA, QDA and KNN. Use at least 3 evaluation metrics.
- For both datasets, compare the performance of the model without interactions and with interactions.

Additional remarks:

- The projects are implemented in teams of 2 students.
- The projects can be implemented in R , Python or other languages.
- The final grade will be based on:
    - source codes (quality of code and structure of project),
    - methodology proposed in part 2,
    - reports (max 4 pages A4) summarizing experiments,
    - presentations (max 5 minutes).
- Please send a zip file (name of the file: surname1_surname2.zip) including 3 folders: code, presentation, report. My e-mail address: norbertryciak(at]gmail.com

- Deadline for both groups: April 16 (Sunday) 2023.
- Presentations (at the University): group 1: April 24 2022, group 2: April 17 2022.
- If you have any questions, please send my an e-mail: norbertryciak(at]gmail.com
- Please let me know the day before (on Sunday) via e-mail if you want to come for a consultation or meet on-line (on Monday morning; other days are also possible).