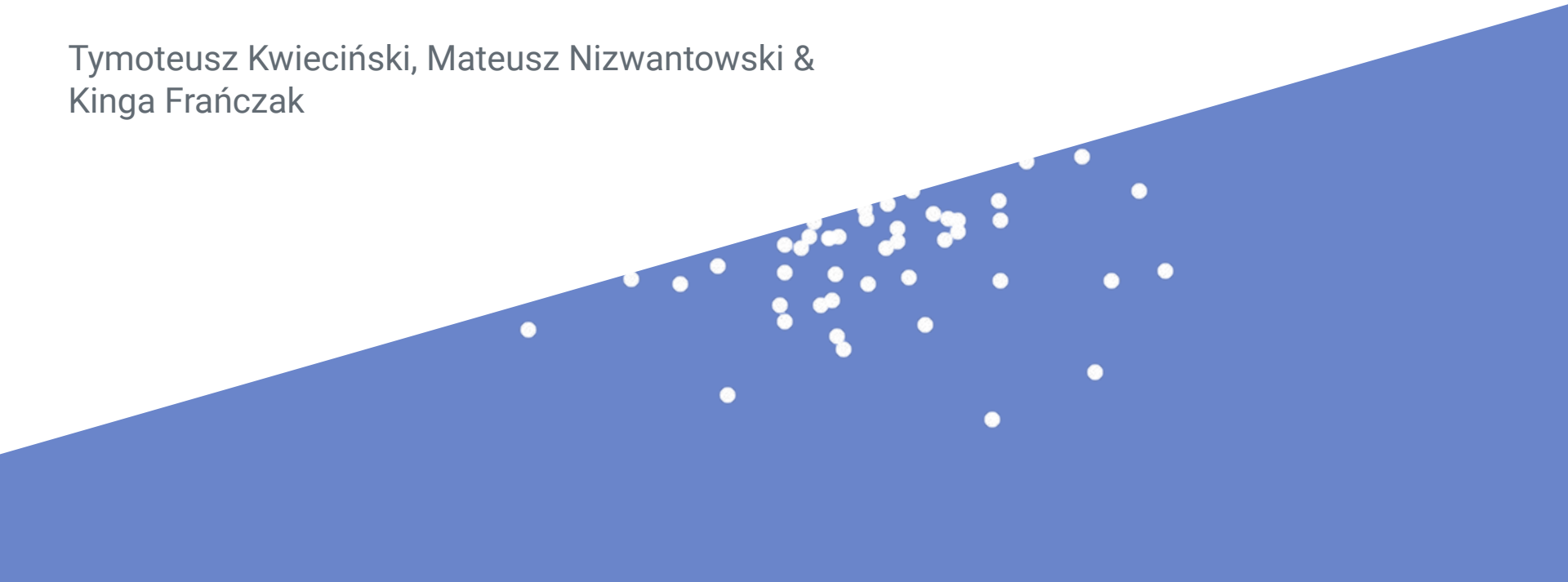
Comparison of different optimization algorithms for **logistic regression**

Tymoteusz Kwieciński, Mateusz Nizwantowski &
Kinga Frańczak



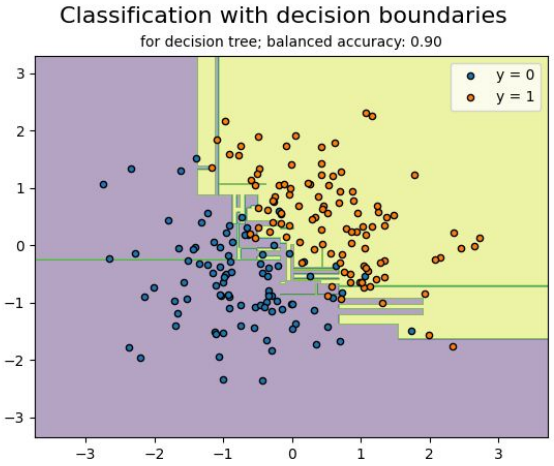
Compared algorithms and methods

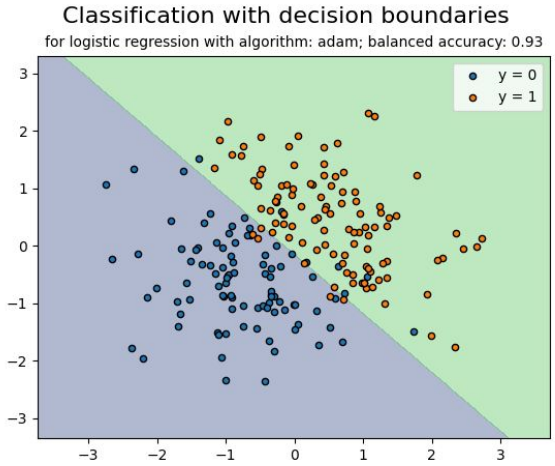
Logistic regression:

1. SGD - Stochastic Gradient Descent
2. Adam - Adaptive Moment Estimation
3. IWLS - Iterative Reweighted Least Squares

Different models:

1. LDA
2. QDA
3. Decision Tree
4. Random Forest





Different behaviour of classifiers on artificially created data

Implementation

We implemented our solution in Python

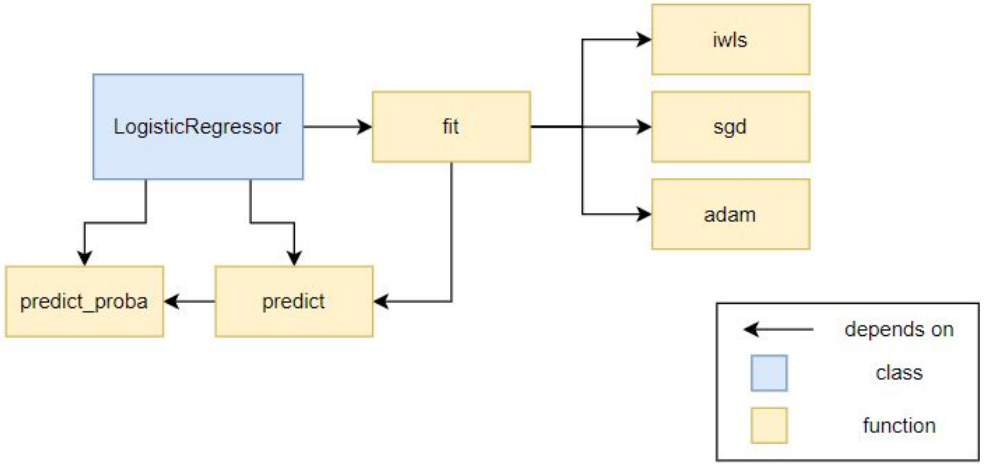
Our code was organized into a simple structure with a single class

Implementation of optimization algorithms was independent from the logistic regression model

For evaluation we used the default parameters for each optimizer

LDA, QDA, Decision Tree and Random Forest were taken from `scikit-learn` library

Brief structure of our implementation of logistic regression model



Datasets

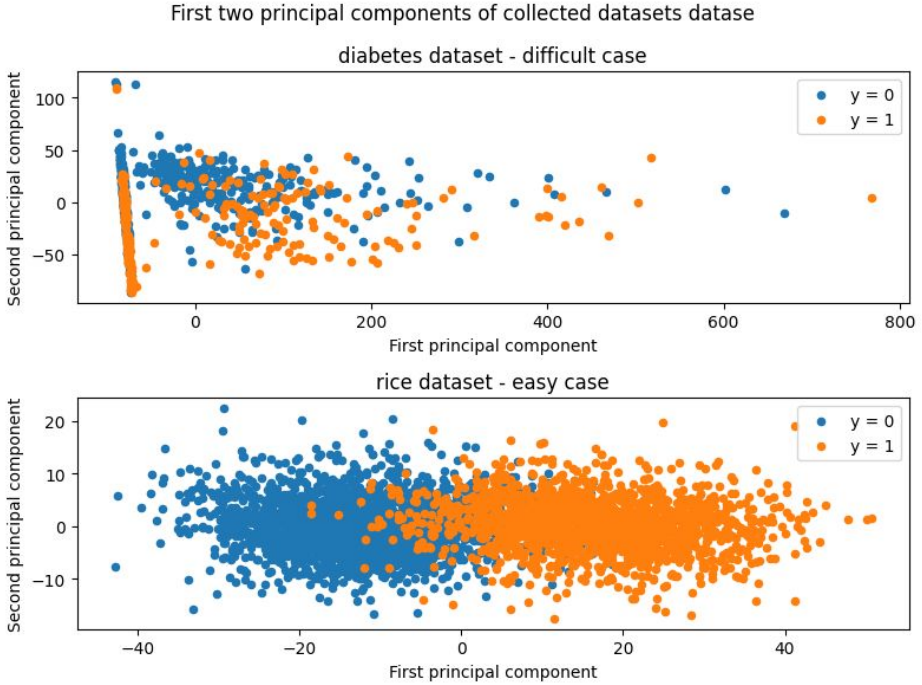
Small datasets (less than 10 variables)

1. diabetes (8 features)
2. rice (7 features)
3. Raisin binary classification (7 features)

Large datasets (more than 10 variables)

4. communities-and-crime-binary (128 features)
5. PolishCompaniesBancrupcy (64 features)
6. AIDS Clinical Trials (23 features)
7. Heart Disease Health Indicators Dataset (21 features)
8. Customer Personality Analysis (29 features)
9. Dry bean - isn't binary but classes were merged (17 features)

Additionally we created an artificial dataset for initial tests and to check if our methods and algorithms work



Evaluation

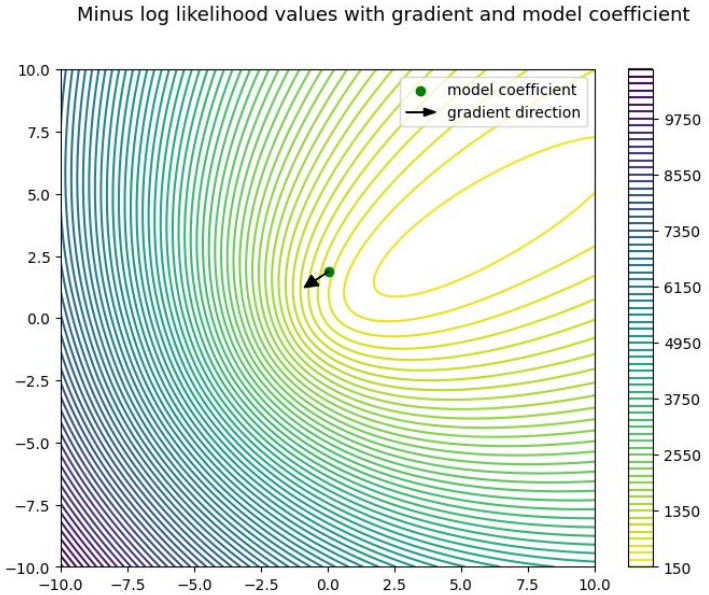
We conducted our evaluation on the collected datasets

To reduce random bias we performed 10 different train-test splits on each dataset to reduce random bias

We agreed on stopping rule which included two restrictions:

1. restriction of gradient L-infinity norm
2. time limit

we halted the training after 500 epochs



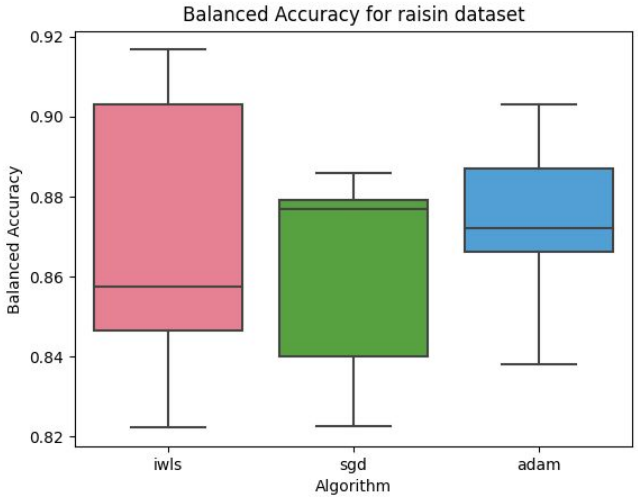
Visualisation of loss in logistic regression model - we wanted to visually inspect that calculated derivatives are correct

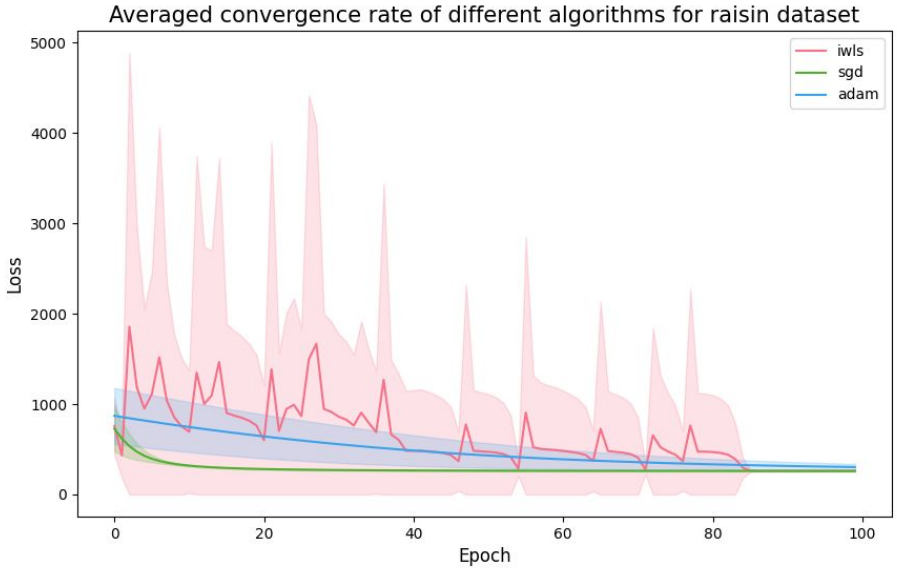
Convergence analysis

SGD was converging in the fastest way

Convergence rate for ADAM was usually stable

IWLS sometimes seemed to don't converge at all





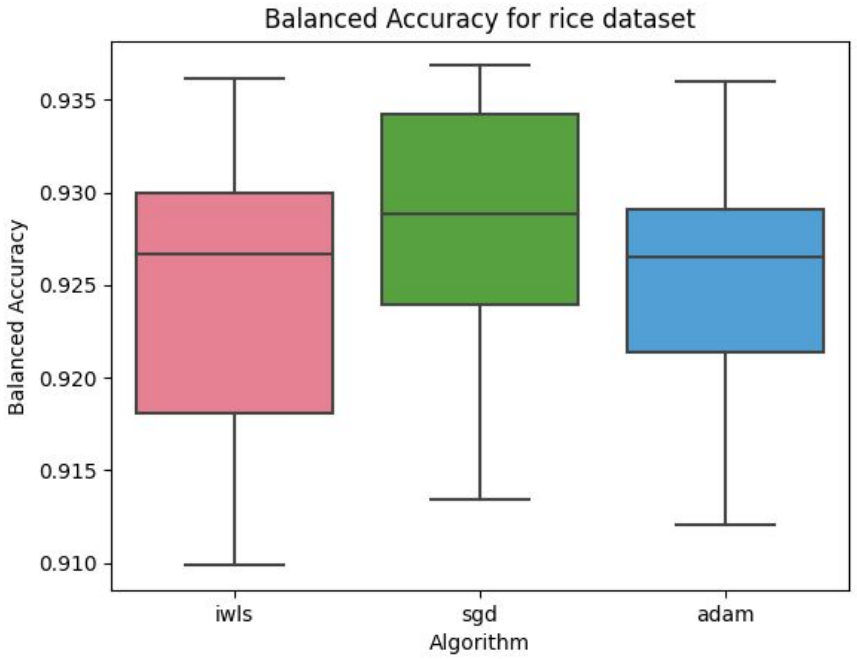
Even though IWLS seems to have troubles with convergence, sometimes it reports better performance on the test set

Classification performance

SGD turned out to be the best when it comes to classification on the test set

ADAM was just bit behind the SGD

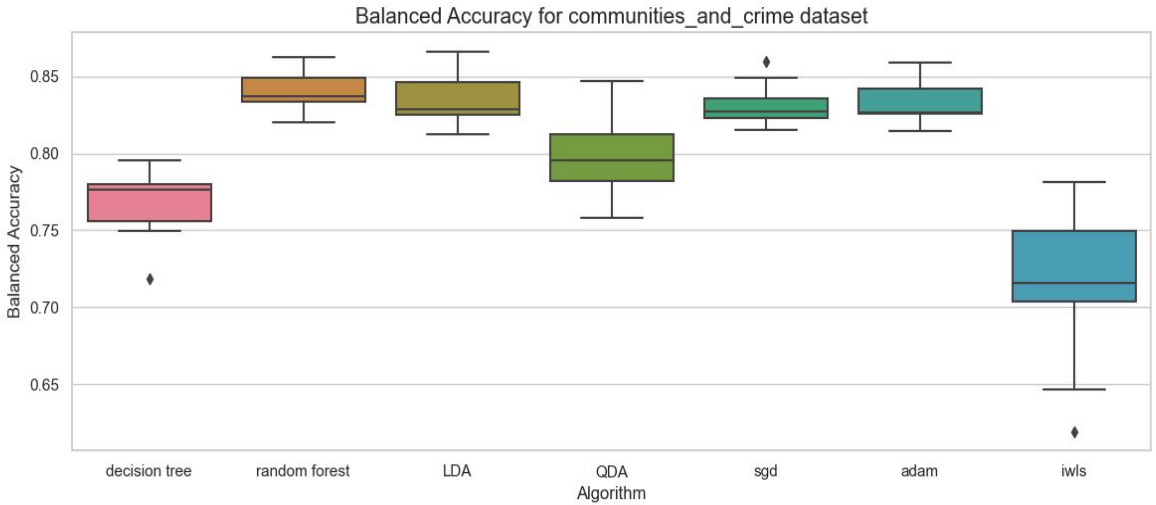
IWLS behaved in the most diverse way; in general had worse results with larger spread of metrics, especially for larger datasets



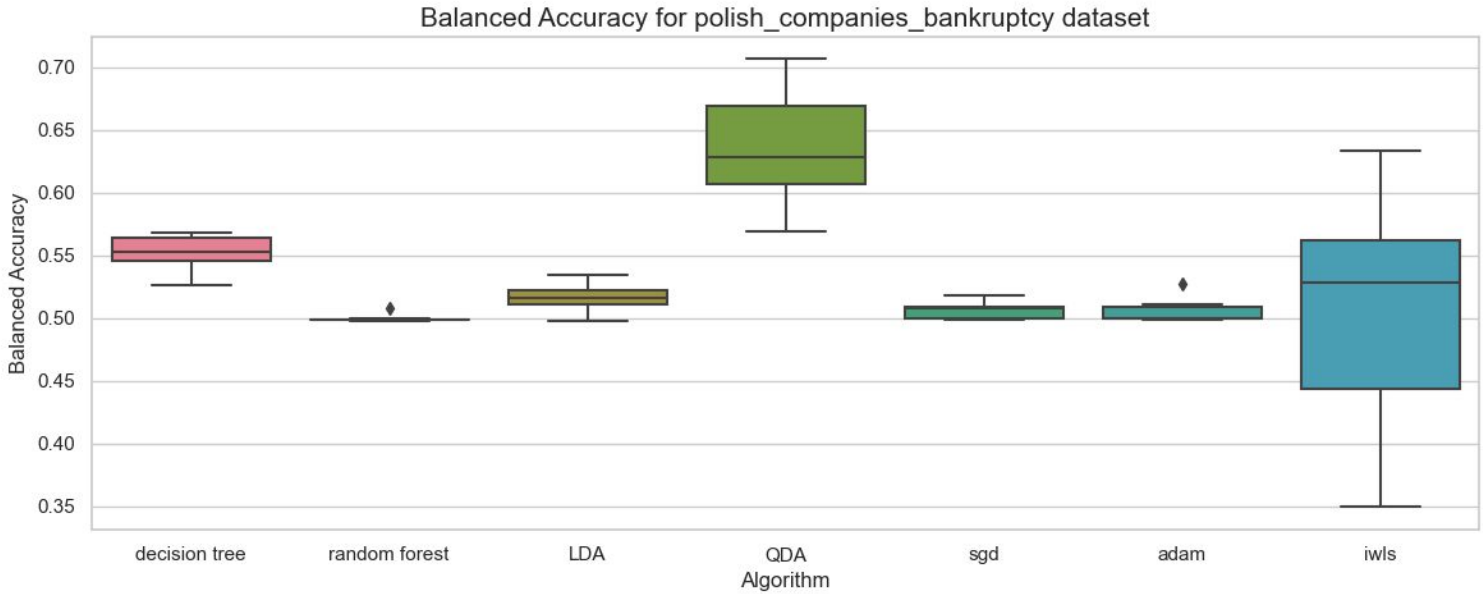
Classification performance in comparison with different algorithms

Logistic regression model turned out to be quite universal and well performing

It didn't report the best results, but they were usually good enough



Classification performance in comparison with different algorithms

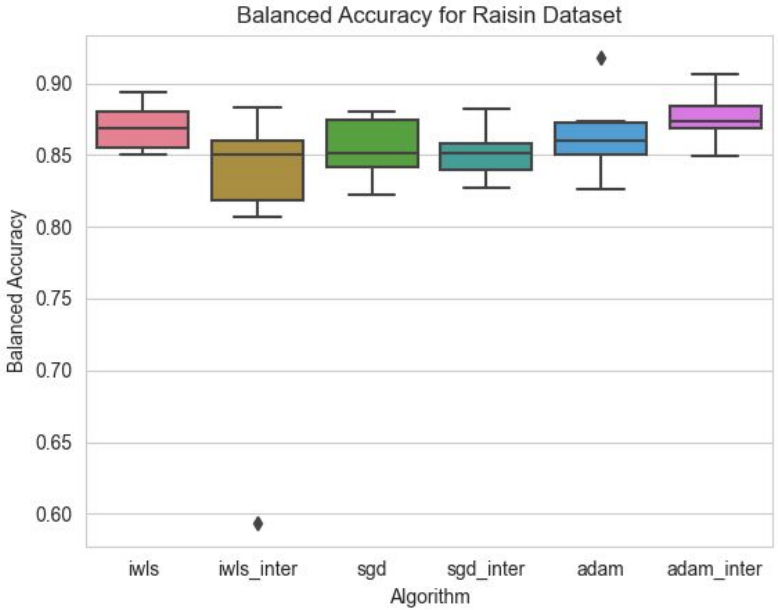


The QDA performs much better than the other methods, logistic regression performance is quite poor; IWLS is again quite unstable, sometimes yielding better results than the remaining optimizers

Classification performance with and without interactions

For all small datasets we tested the performance of our logistic regression model on dataset with and without interactions - multiplications of features

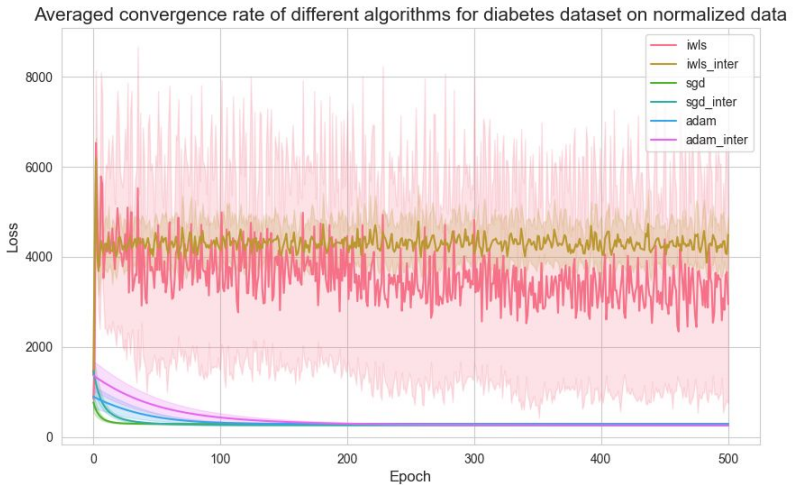
In some cases interactions slightly helped, but in others, probably due to the complexity of the problem, worsened the results

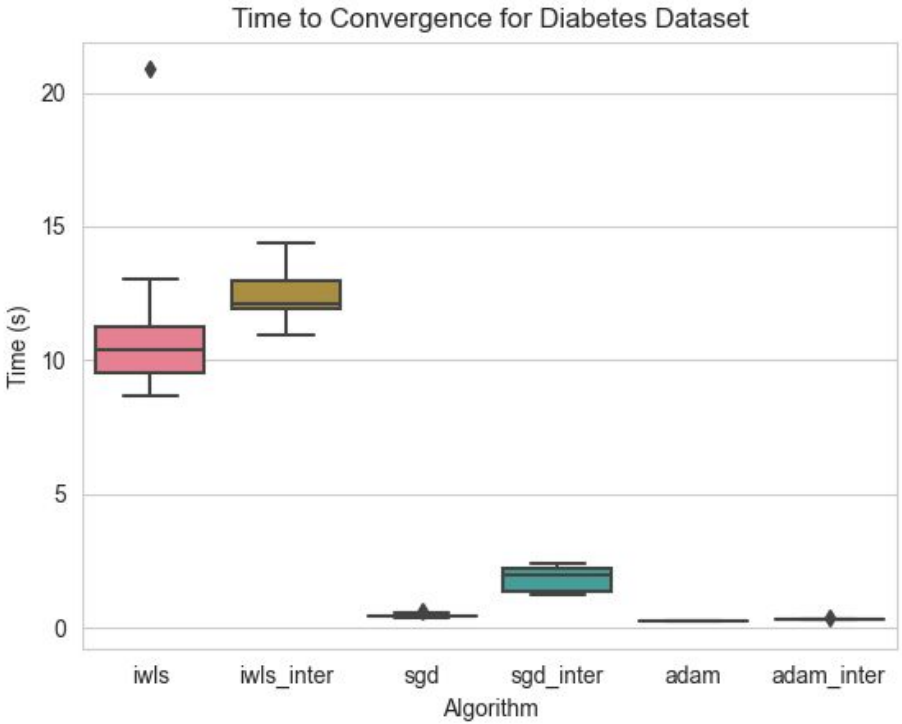


Time measurements

IWLS turned out to be very resource- and time-consuming

In some larger datasets it managed to perform only single one iteration with given constraints





General conclusions

SGD reported the best results, but Adam seems to be only slightly worse in terms of predictive power

Adam is definitely the fastest method, whilst IWLS reports the worst time performance

IWLS is the most unstable reporting sometimes both best and worst results

There is no a single best algorithm nor method for classification

no free lunch theorem illustration; source: https://nmarkou.blogspot.com/2018/04/the-no-free-lunch-theorem.html

Thank you for your attention!