Optimization Algorithms for Logistic Regression

Izabela Telejko, Bartosz Grabek, Grzegorz Zbrzeżny

Agenda

- Datasets and preprocessing stages,
- Introduction to the stopping rule,
- Results of evaluation of IWLS, SGD, and ADAM for classification,
- Comparison of the custom models with popular scikit-learn models,
- Impact of the interactions on optimization results.

Datasets (1)

Dataset	Size	obs.	feat.	Code Name
Tour & Travels Customer Churn	S	954	7	`churn`
Seeds	S	210	8	`seeds`
Pima Indians Diabetes	S	768	9	`diabetes`
Hotel Booking Cancellation	L	36285	17	`booking`
Water Quality	L	7999	21	`water`
Ionosphere	L	351	35	`ionosphere`
Jungle Chess	L	4559	47	`jungle`
League of Legends	L	26834	58	`challenger`
Sonar (Rock vs. Mine)	L	208	61	`sonar`

Datasets (2)

Dataset	% of Class 0	% of Class 1	Imbalance
Tour & Travels Customer Churn	76.50%	23.50%	High
Seeds	66.66%	33.33%	Medium
Pima Indians Diabetes	65.10%	34.90%	Medium
Hotel Booking Cancellation	67.20%	32.80%	Medium
Water Quality	88.56%	11.44%	High
Ionosphere	64.10%	35.90%	Medium
Jungle Chess	55.34%	44.66%	Low
League of Legends	50.01%	49.99%	None
Sonar (Rock vs. Mine)	53.37%	46.63%	Low

Datasets Preprocesssing (1)

Before train-test-split we applied:

- data loading from multiple formats,
- one-hot-encoding and label-encoding the categorical features,
- Encoding target columns to integers 0, 1 for binary classification,
- [Optional] Addition of the interactions between the features.

Datasets Preprocesssing (2)

After train-test-split we applied:

- removal of multicollinear columns on the whole dataset using Variance Inflation Factor,
- standard scaling,
- missing data imputation,
- addition of a column of ones for the bias term.

Custom Stopping Rule

Stop the algorithm if the difference between the current neg. log-likelihood is higher than the best (smallest) recorded neg. log-likelihood reduced by a fixed ε_2=0.1 for t=5 (tolerance) consecutive iterations.

Algorithm 1 Early stopping rule

if $CurrentLogLike < BestLogLike - \epsilon_2$ **then**
  $BestLogLike \leftarrow CurrentLogLike$
  $BestWeights \leftarrow CurrentWeights$
  $NoChangeCounter \leftarrow 0$
else
  $NoChangeCounter \leftarrow NoChangeCounter + 1$
end if
if $NoChangeCounter > t$ **then**
  early stopping
end if

Classification Performance Analysis

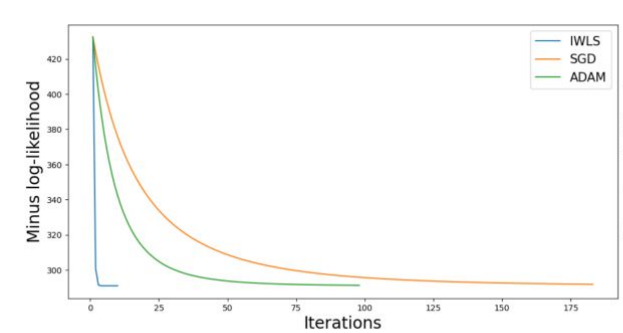
- No significant differences between the methods in terms of final balanced accuracy.
- Worse results are obtained by highly class-imbalanced datasets.

feat.	Code Name	IWLS	SGD	ADAM	Imbalance
7	churn	0.652	0.655	0.659	High
8	seeds	0.970	0.974	0.977	Medium
9	diabetes	0.753	0.755	0.756	Medium
17	booking	0.750	0.752	0.750	Medium
21	water	0.664	0.665	0.663	High
35	ionosphere	0.830	0.835	0.985	Medium
47	jungle	0.986	0.985	0.985	Low
58	challenger	1.000	1.000	1.000	None
61	sonar	0.773	0.780	0.780	Low

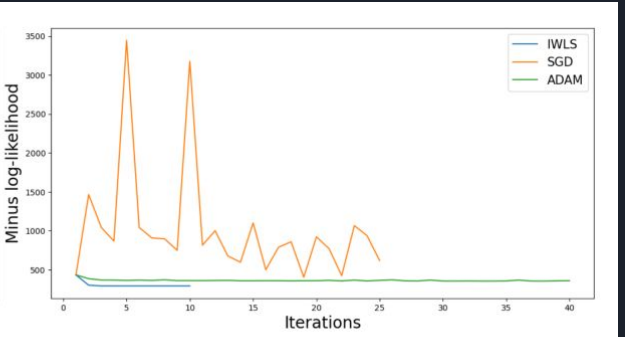
Optimizer Convergence Analysis (1)

Dataset	IWLS	ADAM	SGD
Tour & Travels Customer Churn	10	108	211
Seeds	12	268	501
Pima Indians Diabetes	10	95	185
Hotel Booking Cancellation	13	38	44
Water Quality	12	45	130
Ionosphere	14	479	501
Jungle Chess	16	384	501
League of Legends	18	14	501
Sonar (Rock vs. Mine)	13	424	475

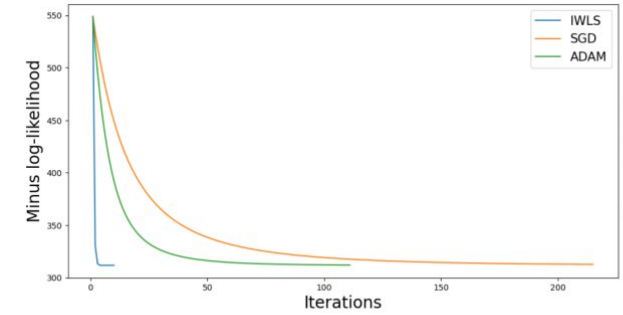
Optimizer Convergence Analysis (2)



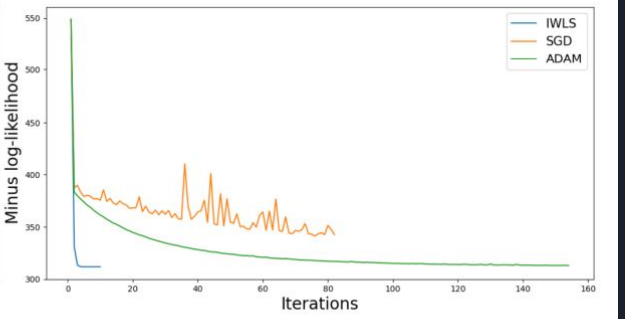
((a)) Diabetes with standard scaling



((b)) Diabetes without scaling

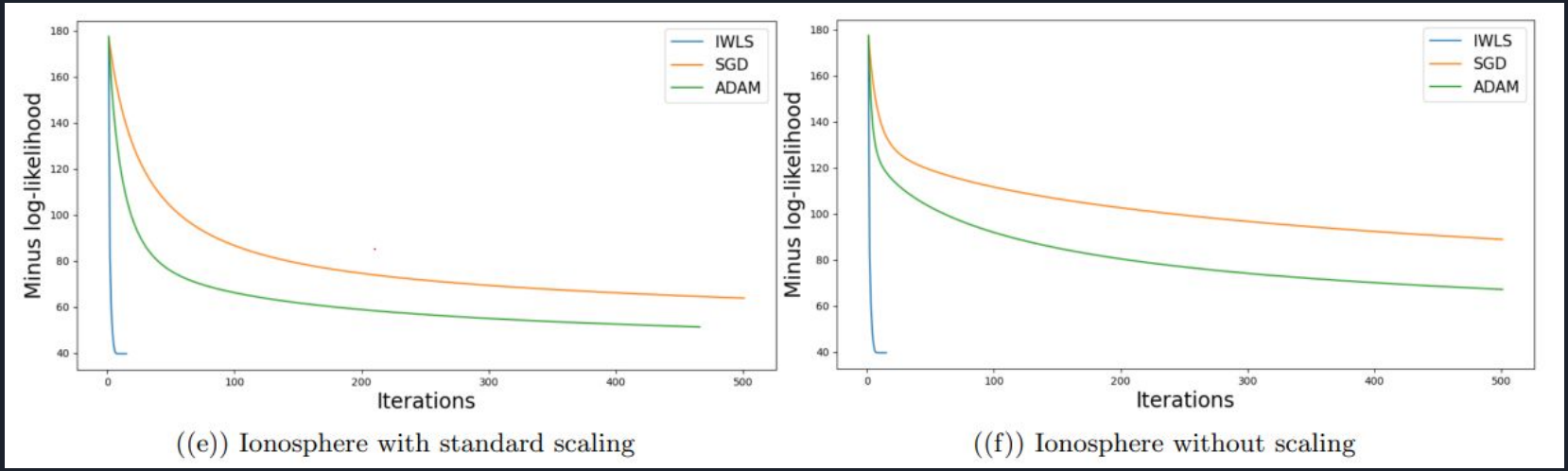


((c)) Churn with standard scaling



((d)) Churn without scaling

Optimizer Convergence Analysis (3)



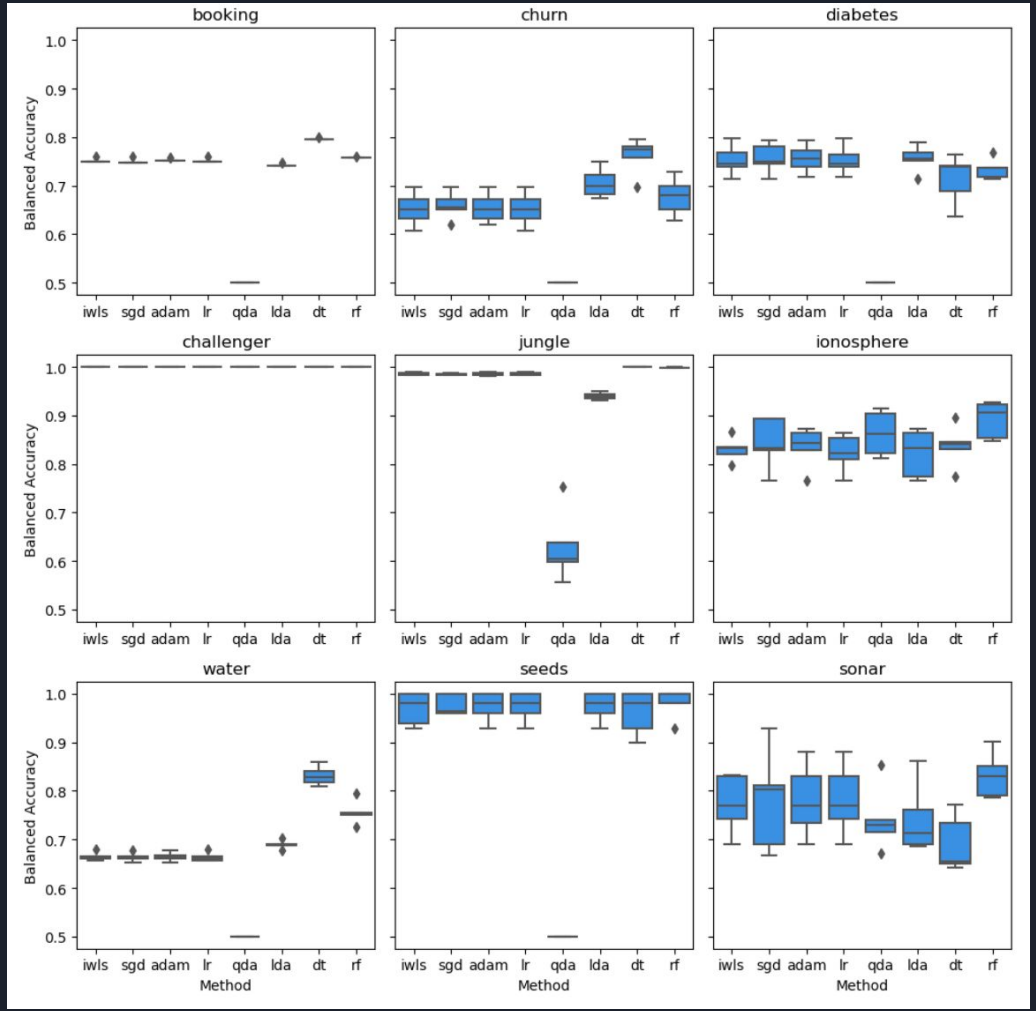
((e)) Ionosphere with standard scaling

((f)) Ionosphere without scaling

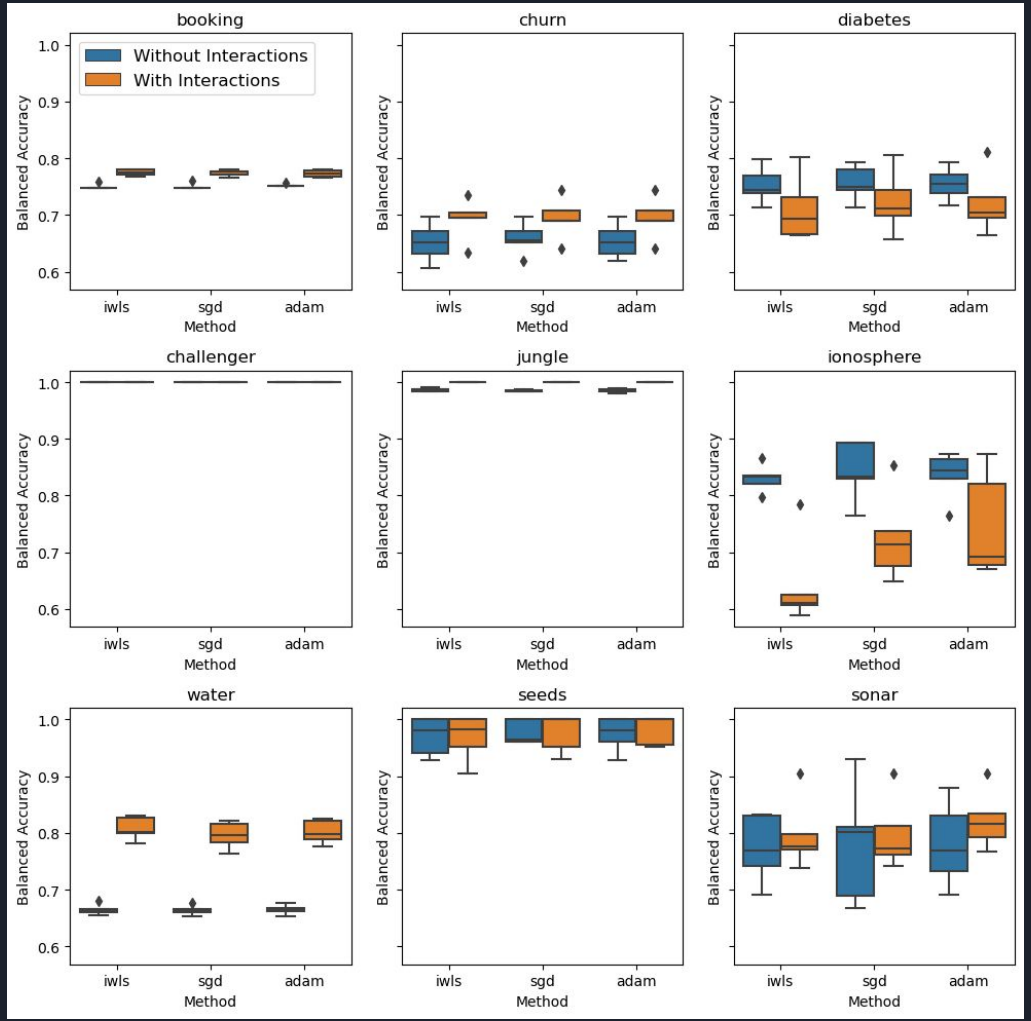
Classification Performance Comparison (1)

Model	Model Name	Setting/Configuration
`iwls`	Iterative Weighted Least Squares (custom)	default
`sgd`	Stochastic Gradient Descent (custom)	`learning_rate=0.0002`
`adam`	Adaptive Moment Estimation (custom)	`learning_rate=0.0002`
`lr`	Logistic Regression	default
`qda`	Quadratic Discriminant Analysis	default
`lda`	Linear Discriminant Analysis	default
`dt`	Decision Tree Classifier	`max_depth=5`
`rf`	Random Forest Classifier	`max_depth=5`, `min_samples_split=3`

Classification Performance Comparison (2)



Classification Performance Comparison on Data with Interactions



Thank you for attention!