OPTIMIZATION ALGORITHMS FOR

LOGISTIC REGRESSION

IWLS | SGD | ADAM

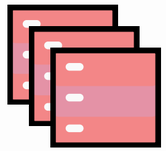
PROJECT 1

ADVANCED MACHINE LEARNING

MATEUSZ BOROWSKI | SZYMON MATUSZEWSKI | MIKOŁAJ ROGUSKI

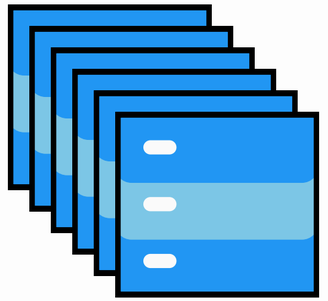
TABLE OF CONTENTS

1. **Datasets**



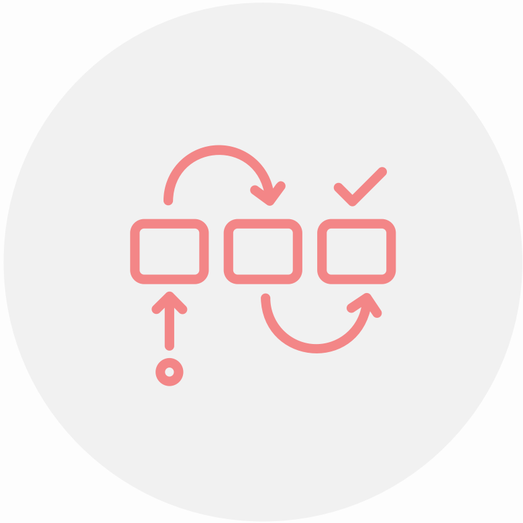
small datasets

name	observations	variables	% of y=1
banknote	1372	4	44
kin8nm	8192	8	49
phoneme	5404	5	29



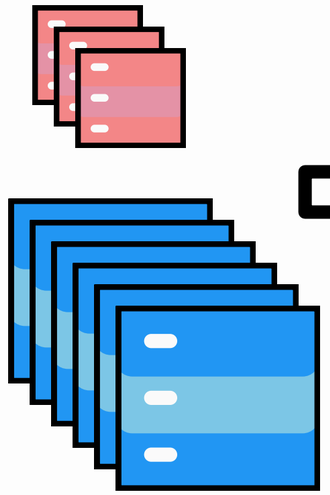
big datasets

name	observations	variables	% of y=1
elevators	16599	18	31
jm1	10885	21	19
ldd_JapaneseVowels	9961	14	16
mfeat_karhunen	2000	64	10
mfeat-zernike	2000	47	10
pc1	1109	21	7



2. Pipeline

datasets



Bigger class -> 0
Smaller class -> 1

Remove
NaNs

Products?

YES

Add
Products

NO

Remove
collinear
features

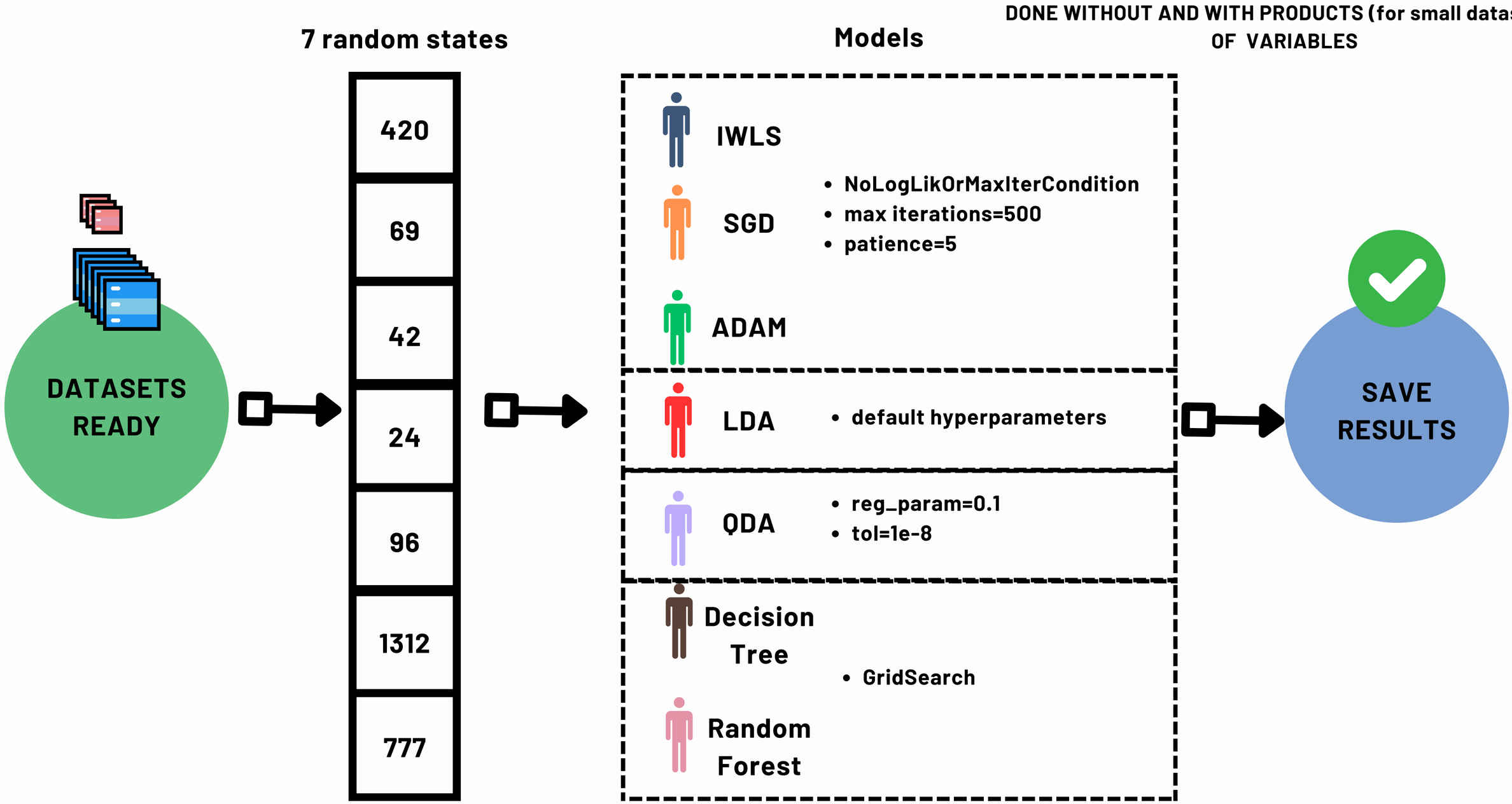
Remove if
more than
85% are 0

Train - Test split

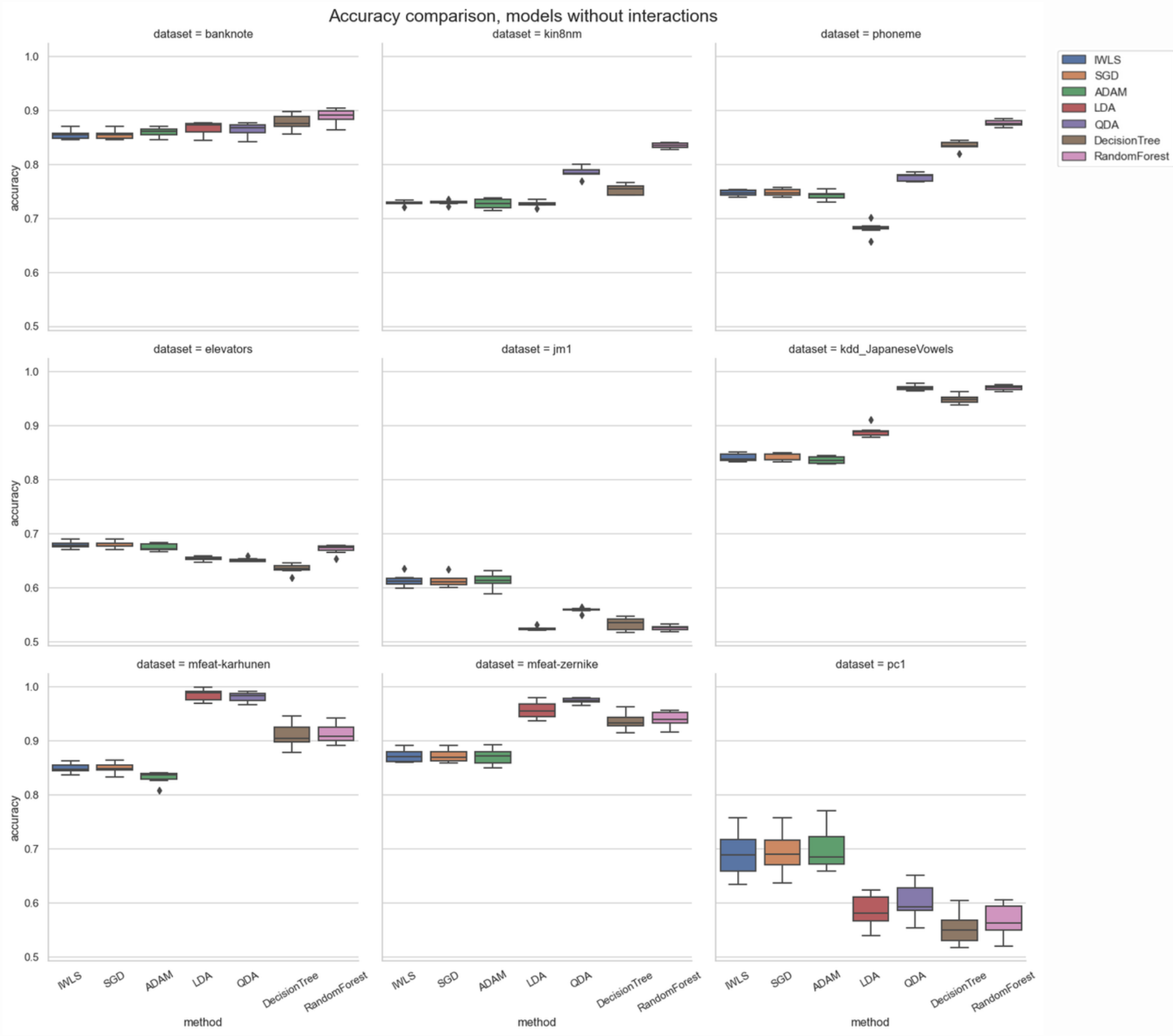
Standard
Scaler

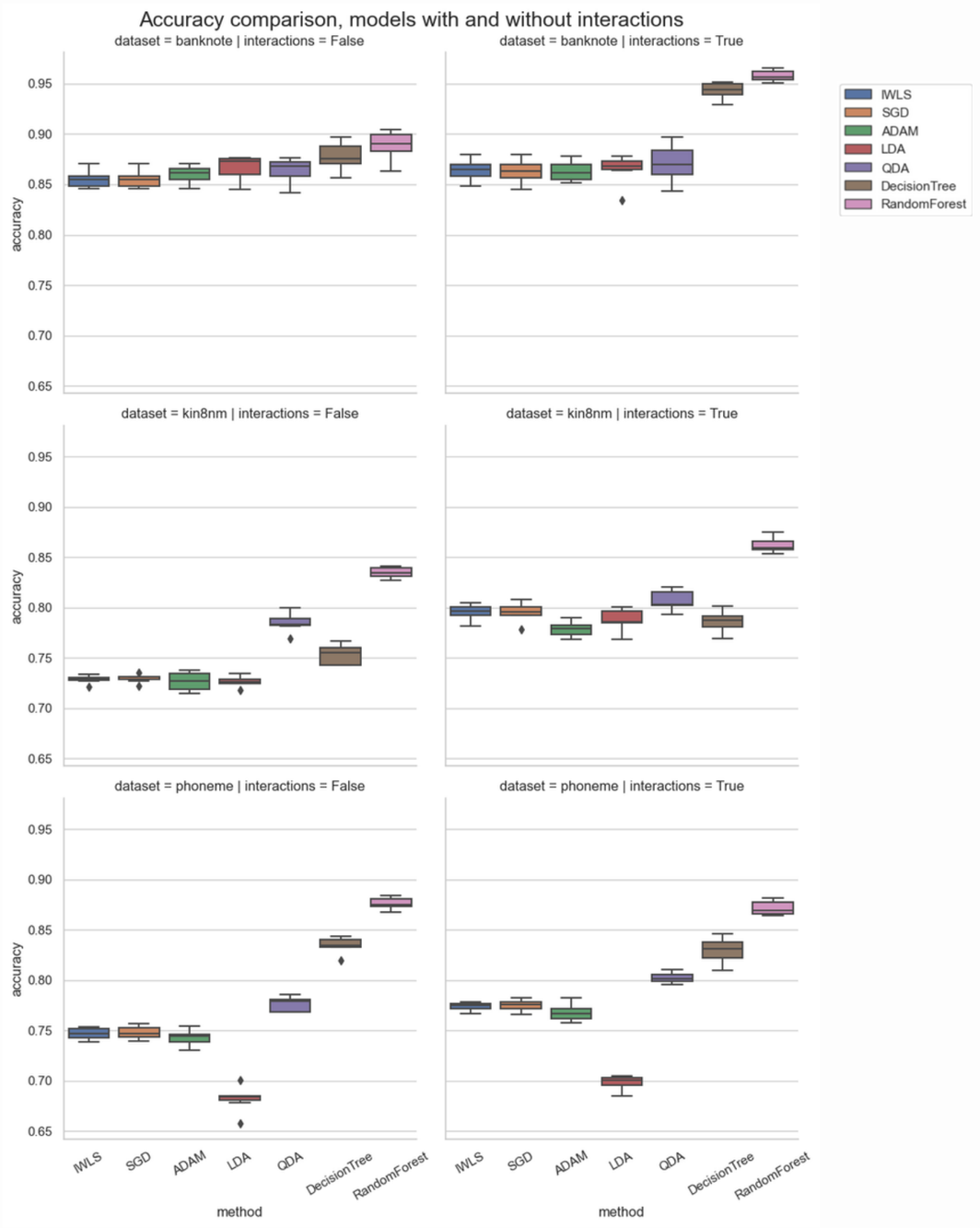
DATASETS
READY

3. Experiments


DONE WITHOUT AND WITH PRODUCTS (for small datasets)
OF VARIABLES
7 random states
Models
DATASETS READY
420
69
42
24
96
1312
777
IWLS
SGD
NoLogLikOrMaxIterCondition
max iterations=500
patience=5
ADAM
LDA
default hyperparameters
QDA
reg_param=0.1
tol=1e-8
Decision Tree
GridSearch
Random Forest
SAVE RESULTS

4. Results


Accuracy comparison, models without interactions
dataset = banknote
dataset = kin8nm
dataset = phoneme
dataset = elevators
dataset = jm1
dataset = kdd_JapaneseVowels
dataset = mfeat-karhunen
dataset = mfeat-zernike
dataset = pc1
accuracy
1.0
0.9
0.8
0.7
0.6
0.5
IWLS
SGD
ADAM
LDA
QDA
DecisionTree
RandomForest
method


Accuracy comparison, models with and without interactions
dataset = banknote | interactions = False
dataset = banknote | interactions = True
dataset = kin8nm | interactions = False
dataset = kin8nm | interactions = True
dataset = phoneme | interactions = False
dataset = phoneme | interactions = True
accuracy
0.95
0.90
0.85
0.80
0.75
0.70
0.65
IWLS
SGD
ADAM
LDA
QDA
DecisionTree
RandomForest
method

BEST 

with products

name	best model	mean acc	std acc
banknote	RF	0.96	0.006
kin8nm	RF	0.86	0.008
phoneme	RF	0.87	0.007

Table 1: Balanced accuracy and the best models for test datasets with products of variables

without products

name	best model	mean acc	std acc
banknote	RF	0.89	0.015
kin8nm	RF	0.83	0.005
phoneme	RF	0.88	0.006
elevators	IWLS	0.68	0.007
jm1	IWLS	0.61	0.011
ldd JapaneseVowels	QDA	0.97	0.005
mfeat_karhunen	LDA	0.98	0.011
mfeat-zernike	QDA	0.97	0.005
pc1	ADAM	0.70	0.041

Table 2: Balanced accuracy and the best models fortest datasets without products of

with products

name	worst model	mean acc	std acc
banknote	ADAM	0.86	0.010
kin8nm	ADAM	0.78	0.007
phoneme	LDA	0.70	0.007

Table 3: Balanced accuracy and the worst models for test datasets with products of variables

without products

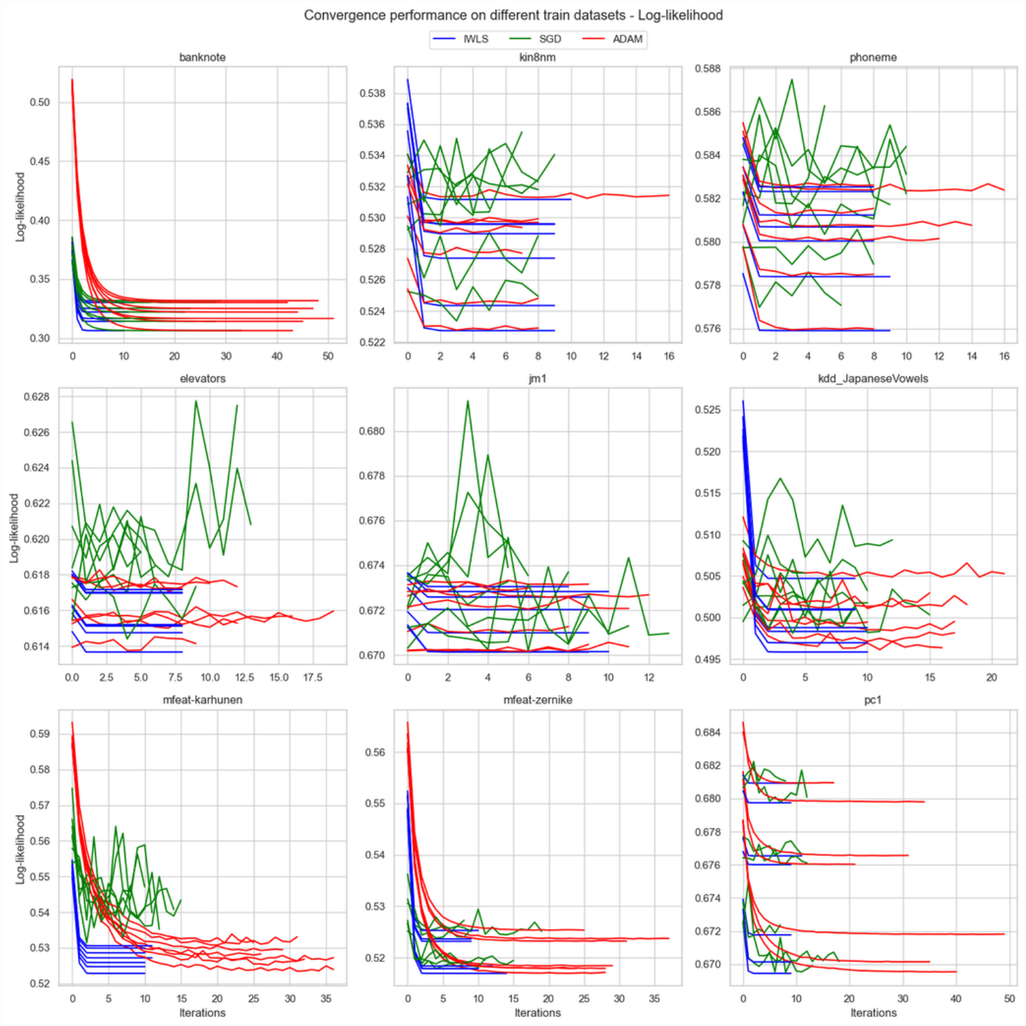
name	worst model	mean acc	std acc
banknote	SGD	0.85	0.009
kin8nm	LDA	0.73	0.005
phoneme	LDA	0.68	0.013
elevators	Decision Tree	0.64	0.009
jm1	LDA	0.52	0.004
ldd JapaneseVowels	ADAM	0.84	0.007
mfeat_karhunen	ADAM	0.83	0.012
mfeat-zernike	ADAM	0.87	0.016
pc1	Decision Tree	0.55	0.031

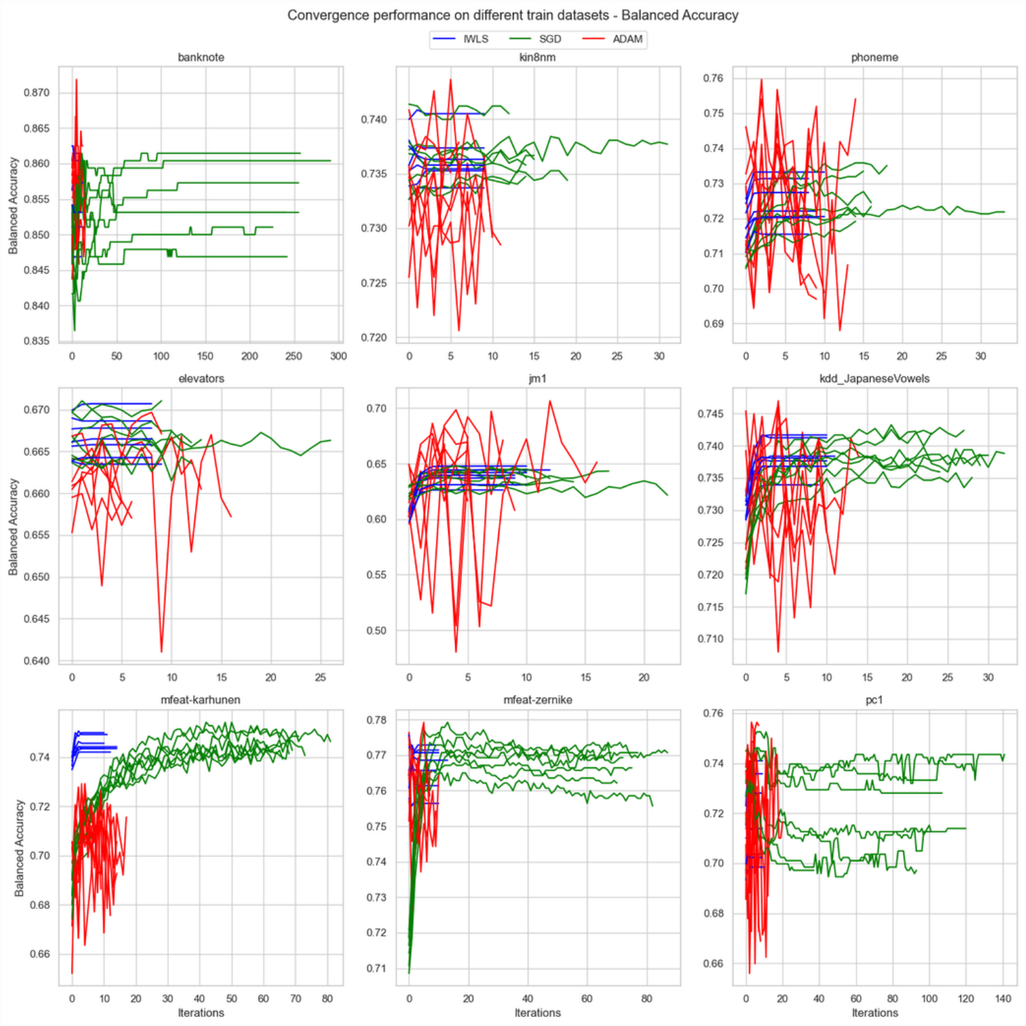
Table 4: Balanced accuracy and the worst models for test datasets without products of variables

WHICH OPTIMIZER IS THE BEST?

name	IWLS	ADAM	SGD
banknote	1	3	2
kin8nm	1	3	2
phoneme	2	3	1
elevators	1	3	2
jm1	1	3	2
ldd JapaneseVowels	2	3	1
mfeat_karhunen	2	3	1
mfeat-zernike	1	3	2
pc1	3	1	2

Table 5: Standings within the ranking for the Logistic Regression optimizers based on mean balanced accuracy on test datasets. Scores for small datasets are evaluated by datasets with products.

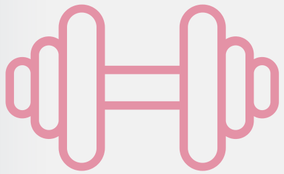

Convergence performance on different train datasets - Log-likelihood
IWLS
SGD
ADAM
banknote
kin8nm
phoneme
elevators
jm1
kdd_JapaneseVowels
mfeat-karhunen
mfeat-zernike
pc1
Log-likelihood
Iterations


Convergence performance on different train datasets - Balanced Accuracy
IWLS
SGD
ADAM
banknote
kin8nm
phoneme
elevators
jm1
kdd_JapaneseVowels
mfeat-karhunen
mfeat-zernike
pc1
Balanced Accuracy
Iterations

5. Conclusions

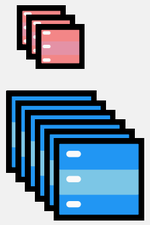
FINAL CONCLUSIONS

Random Forests



IWLS

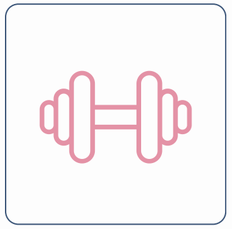
ADAM



Small vs Big Datasets

Random Forests

It is easy to conclude that for most cases the Random Forest algorithm achieves the best balanced accuracy outperforming even the Logistic Regression optimizers like IWLS, ADAM and SGD

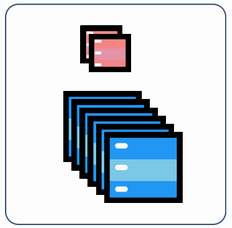


IWLS best optimizer for Logistic Regression

Comparing only optimizers' performance it is IWLS algorithm which achieves the best scores. It was the best for 2 of the big datasets and its convergence seems to be the fastest and controlable.

ADAM is outperformed by other optimizers

ADAM optimizer seems to be outperformed in terms of balanced accuracy by other optimizers but it is possibly because of a non-tuned learning rate.



Small vs Big Datasets

For small datasets the simplest algortihms are the best in terms of balanced accuracy, i.e. Random Forests. Bigger require different approaches like IWLS or QDA.


Thank
you!