Contents

1 new experiments: chemReg with cAdam

1.1 new experiments:chemReg with cAdam sorted by val MAE

parameter name	best values				
'final val loss mae avg' (avg)	0.1116	0.11208	0.11416	0.11493	0.11534
test loss mae avg	0.11389	0.11298	0.11493	0.11445	0.11539
test loss avg	0.012971	0.012764	0.0065903	0.0065354	0.0066426
training time avg	44.379	566.85	577.92	80.447	44.17
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(30, 30, 10)	(40, 20)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	MSE	log cosh	log cosh	log cosh
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	1000	100	100	1000	1000
optimizer	Adam	cAdam	cAdam	cAdam	Adam
learning rate	0.01	0.001	0.001	0.01	0.01
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 1.1: best settings regarding *final val loss mae avg* for the chemReg cAdam dataset

parameter name	parameter values			win ratios in %			avg. differences			best value
neurons per layer	(40, 20)	(50, 10)	(30, 30, 10)	29.2	50.0	20.8	0.005	0.003	0.007	(50, 10)
activation functions	ReLU	sigmoid		50.0	50.0		0.003	0.006		ReLU
loss function	MSE	log cosh		31.9	68.1		0.004	0.002		log cosh
batch size	100	1000	10000	52.1	45.8	2.1	0.005	0.005	0.032	*unclear*
optimizer	Adam	cAdam		41.7	58.3		0.005	0.002		cAdam
learning rate	0.01	0.001		63.9	36.1		0.004	0.015		0.01

Table 1.2: parameter influence regarding *final val loss mae avg* for the chemReg cAdam dataset

parameter name	worst values				
'final val loss mae avg' (avg)	0.19109	0.19588	0.20277	0.21157	0.22639
test loss mae avg	0.18477	0.19365	0.19938	0.20834	0.22996
test loss avg	0.034141	0.018634	0.019747	0.043405	0.052883
training time avg	17.242	17.453	17.19	17.624	16.914
neurons per layer	(50, 10)	(50, 10)	(40, 20)	(30, 30, 10)	(40, 20)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	log cosh	log cosh	MSE	MSE
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	10000	10000	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	Adam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 1.3: worst settings regarding *final val loss mae avg* for the chemReg cAdam dataset

2 MNIST Bachelorthesis new result tables

2.1 accuracy

parameter name	best values				
final val accuracy avg	0.97	0.9691	0.96905	0.96885	0.9688
test accuracy avg	0.97006	0.96906	0.9681	0.97048	0.97084
training time avg	46.666	43.602	86.105	86.874	23.122
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU	ReLU
last activation function	sigmoid	sigmoid	sigmoid	softmax	softmax
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1	1	1	1	1
number of epochs	50	25	50	50	25
batch size	100	100	100	100	100

optimizer	Adam	cAdam	cAdam	cAdam	Adam
learning rate	0.001	0.1	0.1	0.1	0.1
ε	10^{-7}	1.0	1.0	1.0	1.0

Table 2.1: best settings regarding *final val accuracy avg* for the MNIST BSc dataset

parameter name	**worst values**				
final val accuracy avg	0.085333	0.0852	0.085183	0.084467	0.0833
test accuracy avg	0.0845	0.0821	0.08686	0.08748	0.08516
training time avg	1.4319	6.3722	3.1055	3.1968	1.3563
neurons per layer	(32,)	(20, 15, 10)	(32,)	(32,)	(50, 10)
activation functions	ReLU	ReLU	ReLU	sigmoid	ReLU
last activation function	softmax	sigmoid	sigmoid	sigmoid	sigmoid
loss function	MSE	MSE	MSE	cat-cross	MSE
training data percentage	1	1	1	1	1
number of epochs	5	5	25	25	5
batch size	1000	100	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	cAdam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	1.0	1.0	1.0	1.0	1.0

Table 2.2: worst settings regarding *final val accuracy avg* for the MNIST BSc dataset

parameter name	**parameter values**			**win ratios in %**			**avg. differences**			**best value**
neurons per layer	(32,)	(50, 10)	(20, 15, 10)	63.8	31.4	4.9	0.003	0.117	0.17	(32,)
activation functions	ReLU		sigmoid	73.1		26.9	0.062		0.114	ReLU
last activation function	softmax		sigmoid	62.3		37.7	0.007		0.069	softmax
loss function	cat-cross		MSE	86.7		13.3	0.001		0.221	cat-cross
number of epochs	5	25	50	6.8	19.2	74.0	0.159	0.047	0.006	50
batch size	100	1000	10000	74.5	19.4	6.0	0.023	0.131	0.283	100
optimizer	Adam		cAdam	34.6		65.4	0.031		0.01	cAdam
learning rate	0.1	0.01	0.001	46.9	35.1	18.1	0.175	0.096	0.202	*unclear*
ε	1.0		10^{-7}	16.2		83.8	0.437		0.058	10^{-7}

Table 2.3: parameter influence regarding *final val accuracy avg* for the MNIST BSc dataset

3 MNIST Adam variant comparison

3.1 revised, longer experiments, 1 run

3.2 training time

parameter name	best values			
training time avg	55.489	62.251	92.568	96.35
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU
last activation function	sigmoid	softmax	sigmoid	softmax
loss function	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1.0	1.0	1.0	1.0
number of epochs	250	250	250	250
batch size	100	100	100	100
optimizer	Adam	Adam	my Adam	my Adam
learning rate	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 3.1: best settings regarding *training time avg* for the cadam variants dataset

parameter name	worst values				
training time avg	96.35	104.57	110.17	113.87	115.73
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU	ReLU
last activation function	softmax	softmax	sigmoid	softmax	sigmoid
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1.0	1.0	1.0	1.0	1.0
number of epochs	250	250	250	250	250
batch size	100	100	100	100	100

optimizer	my Adam	c adam hat	c adam hat	cAdam	cAdam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 3.2: worst settings regarding *training time avg* for the cadam variants dataset

parameter name	**parameter values**						**win ratios in %**						**avg. differences in %**		**best value**
last activation function	softmax			sigmoid			50.0			50.0			4.068	1.747	sigmoid
optimizer	Adam	my Adam	cAdam	cAdam$_hat$	100.0	0	0	0	0	60.799	95.747	83.259	Adam		

Table 3.3: parameter influence regarding *training time avg* for the cadam variants dataset

3.3 accuracy

parameter name	**best values**				
test accuracy avg	0.1943	0.1941	0.19394	0.1939	0.1938
final val accuracy avg	0.19403	0.19362	0.19413	0.19363	0.1939
final val accuracy std	0.38807	0.38723	0.38827	0.38727	0.3878
final val accuracy min	0.0	0.0	0.0	0.0	0.0
final val accuracy max	0.97017	0.96808	0.97067	0.96817	0.9695
training time avg	96.35	113.87	92.568	55.489	115.73
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU	ReLU
last activation function	softmax	softmax	sigmoid	sigmoid	sigmoid
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1.0	1.0	1.0	1.0	1.0
number of epochs	250	250	250	250	250
batch size	100	100	100	100	100
optimizer	my Adam	cAdam	my Adam	Adam	cAdam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 3.4: best settings regarding *test accuracy avg* for the cadam variants dataset

parameter name	worst values				
test accuracy avg	0.1939	0.1938	0.19372	0.13374	0.13306
final val accuracy avg	0.19363	0.1939	0.1939	0.13457	0.13483
final val accuracy std	0.38727	0.3878	0.3878	0.26913	0.26967
final val accuracy min	0.0	0.0	0.0	0.0	0.0
final val accuracy max	0.96817	0.9695	0.9695	0.67283	0.67417
training time avg	55.489	115.73	62.251	110.17	104.57
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU	ReLU
last activation function	sigmoid	sigmoid	softmax	sigmoid	softmax
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1.0	1.0	1.0	1.0	1.0
number of epochs	250	250	250	250	250
batch size	100	100	100	100	100
optimizer	Adam	cAdam	Adam	c adam hat	c adam hat
learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 3.5: worst settings regarding *test accuracy avg* for the cadam variants dataset

parameter name	parameter values						win ratios in %						avg. differences in %		best value
last activation function	softmax			sigmoid			50.0			50.0			0.15	0.085	sigmoid
optimizer	Adam	my Adam	cAdam	cAdam$_hat$	0	100.0	0	0	0.16	0	0.088	31.279	my Adam		

Table 3.6: parameter influence regarding *test accuracy avg* for the cadam variants dataset

4 MNIST with % differences

4.1 training time

parameter name	best values				
training time avg	1.0347	1.0368	1.0398	1.0412	1.0446
neurons per layer	(32,)	(32,)	(32,)	(32,)	(32,)

activation functions	sigmoid	sigmoid	ReLU	ReLU	sigmoid
last activation function	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
loss function	MSE	MSE	MSE	MSE	MSE
training data percentage	1	1	1	1	1
number of epochs	5	5	5	5	5
batch size	10000	10000	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	Adam
learning rate	0.001	0.01	0.001	0.1	0.1
ε	10^{-7}	10^{-7}	10^{-7}	1.0	1.0

Table 4.1: best settings regarding *training time avg* for the MNIST revised dataset

parameter name	**worst values**				
training time avg	106.16	106.61	106.64	106.92	106.93
neurons per layer	(20, 15, 10)	(20, 15, 10)	(20, 15, 10)	(20, 15, 10)	(20, 15, 10)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	softmax	softmax	softmax	softmax	softmax
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1	1	1	1	1
number of epochs	50	50	50	50	50
batch size	100	100	100	100	100
optimizer	cAdam	cAdam	cAdam	cAdam	cAdam
learning rate	0.001	0.01	0.1	0.01	0.1
ε	1.0	1.0	1.0	10^{-7}	10^{-7}

Table 4.2: worst settings regarding *training time avg* for the MNIST revised dataset

parameter name	**parameter values**			**win ratios in %**			**avg. differences in %**			**best value**
neurons per layer	(32,)	(50, 10)	(20, 15, 10)	98.6	1.2	0.2	0.036	12.017	22.725	(32,)
activation functions	ReLU		sigmoid	49.9		50.1	0.716		0.602	sigmoid
last activation function	softmax		sigmoid	24.8		75.2	1.296		0.506	sigmoid
loss function	cat-cross		MSE	52.6		47.4	1.241		4.097	cat-cross
number of epochs	5	25	50	100.0	0	0	0	264.684	596.88	5

batch size	100	1000	10000	0	0	100.0	779.878	64.515	0	10000
optimizer	Adam		cAdam	100.0	0		0	39.945		Adam
learning rate	0.1	0.01	0.001	34.7	30.9	34.4	0.914	1.126	1.094	0.1
ε	1.0		10^{-7}	54.9	45.1		0.666	0.686		1.0

Table 4.3: parameter influence regarding *training time avg* for the MNIST revised dataset

4.2 accuracy

parameter name	**best values**				
final val accuracy avg	0.97	0.9691	0.96905	0.96885	0.9688
test accuracy avg	0.97006	0.96906	0.9681	0.97048	0.97084
training time avg	46.666	43.602	86.105	86.874	23.122
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(50, 10)	(50, 10)
activation functions	ReLU	ReLU	ReLU	ReLU	ReLU
last activation function	sigmoid	sigmoid	sigmoid	softmax	softmax
loss function	cat-cross	cat-cross	cat-cross	cat-cross	cat-cross
training data percentage	1	1	1	1	1
number of epochs	50	25	50	50	25
batch size	100	100	100	100	100
optimizer	Adam	cAdam	cAdam	cAdam	Adam
learning rate	0.001	0.1	0.1	0.1	0.1
ε	10^{-7}	1.0	1.0	1.0	1.0

Table 4.4: best settings regarding *final val accuracy avg* for the MNIST revised dataset

parameter name	**worst values**				
final val accuracy avg	0.085333	0.0852	0.085183	0.084467	0.0833
test accuracy avg	0.0845	0.0821	0.08686	0.08748	0.08516
training time avg	1.4319	6.3722	3.1055	3.1968	1.3563
neurons per layer	(32,)	(20, 15, 10)	(32,)	(32,)	(50, 10)
activation functions	ReLU	ReLU	ReLU	sigmoid	ReLU
last activation function	softmax	sigmoid	sigmoid	sigmoid	sigmoid

loss function	MSE	MSE	MSE	cat-cross	MSE
training data percentage	1	1	1	1	1
number of epochs	5	5	25	25	5
batch size	1000	100	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	cAdam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	1.0	1.0	1.0	1.0	1.0

Table 4.5: worst settings regarding *final val accuracy avg* for the MNIST revised dataset

parameter name	**parameter values**			**win ratios in %**			**avg. differences in %**			**best value**
neurons per layer	(32,)	(50, 10)	(20, 15, 10)	63.8	31.4	4.9	1.091	20.613	27.614	(32,)
activation functions	ReLU		sigmoid	73.1	26.9		8.66	20.247		ReLU
last activation function	softmax		sigmoid	62.3	37.7		2.423	10.791		softmax
loss function	cat-cross		MSE	86.7	13.3		0.401	33.215		cat-cross
number of epochs	5	25	50	6.8	19.2	74.0	25.91	9.915	1.916	50
batch size	100	1000	10000	74.5	19.4	6.0	4.087	21.393	38.937	100
optimizer	Adam		cAdam	34.6	65.4		7.61	2.37		cAdam
learning rate	0.1	0.01	0.001	46.9	35.1	18.1	20.242	16.717	30.161	*unclear*
ε	1.0		10^{-7}	16.2	83.8		52.679	7.421		10^{-7}

Table 4.6: parameter influence regarding *final val accuracy avg* for the MNIST revised dataset

5 ChemRegB with % differences

5.1 training time

parameter name	**best values**				
training time avg	16.508	16.688	16.872	16.896	16.914
neurons per layer	(40, 20)	(40, 20)	(40, 20)	(40, 20)	(40, 20)
activation functions	ReLU	ReLU	ReLU	ReLU	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	MSE	log cosh	log cosh	MSE

training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	10000	10000	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	Adam
learning rate	0.001	0.01	0.01	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.1: best settings regarding *training time avg* for the chemReg Adam dataset

parameter name	**worst values**				
training time avg	678.56	691.95	692.12	696.0	697.49
neurons per layer	(30, 30, 10)	(30, 30, 10)	(30, 30, 10)	(30, 30, 10)	(30, 30, 10)
activation functions	sigmoid	ReLU	ReLU	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	log cosh	log cosh	log cosh	log cosh
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	100	100	100	100	100
optimizer	cAdam	cAdam	cAdam	cAdam	cAdam
learning rate	0.01	0.01	0.001	0.001	0.01
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.2: worst settings regarding *training time avg* for the chemReg Adam dataset

parameter name	**parameter values**			**win ratios in %**			**avg. differences in %**			**best value**
neurons per layer	(40, 20)	(50, 10)	(30, 30, 10)	87.5	12.5	0	0.151	1.948	11.851	(40, 20)
activation functions	ReLU	sigmoid		51.4	48.6		0.63	0.614		*unclear*
loss function	MSE	log cosh		87.5	12.5		0.126	2.565		MSE
batch size	100	1000	10000	0	0	100.0	2448.099	214.332	0	10000
optimizer	Adam	cAdam		100.0	0		0	47.609		Adam
learning rate	0.01	0.001		43.1	56.9		0.527	0.577		*unclear*

Table 5.3: parameter influence regarding *training time avg* for the chemReg Adam dataset

5.2 validation loss

parameter name	best values				
'final val loss mae avg' (avg)	0.1116	0.11208	0.11416	0.11493	0.11534
test loss mae avg	0.11389	0.11298	0.11493	0.11445	0.11539
test loss avg	0.012971	0.012764	0.0065903	0.0065354	0.0066426
training time avg	44.379	566.85	577.92	80.447	44.17
neurons per layer	(50, 10)	(50, 10)	(50, 10)	(30, 30, 10)	(40, 20)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	MSE	log cosh	log cosh	log cosh
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	1000	100	100	1000	1000
optimizer	Adam	cAdam	cAdam	cAdam	Adam
learning rate	0.01	0.001	0.001	0.01	0.01
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.4: best settings regarding *final val loss mae avg* for the chemReg Adam dataset

parameter name	worst values				
'final val loss mae avg' (avg)	0.19109	0.19588	0.20277	0.21157	0.22639
test loss mae avg	0.18477	0.19365	0.19938	0.20834	0.22996
test loss avg	0.034141	0.018634	0.019747	0.043405	0.052883
training time avg	17.242	17.453	17.19	17.624	16.914
neurons per layer	(50, 10)	(50, 10)	(40, 20)	(30, 30, 10)	(40, 20)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	log cosh	log cosh	MSE	MSE
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	10000	10000	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	Adam

learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.5: worst settings regarding *final val loss mae avg* for the chemReg Adam dataset

parameter name	parameter values			win ratios in %			avg. differences in %			best value
neurons per layer	(40, 20)	(50, 10)	(30, 30, 10)	29.2	50.0	20.8	3.395	2.134	5.689	(50, 10)
activation functions	ReLU	sigmoid		50.0	50.0		2.743	4.037		ReLU
loss function	MSE	log cosh		31.9	68.1		2.961	1.155		log cosh
batch size	100	1000	10000	52.1	45.8	2.1	4.485	4.275	26.929	*unclear*
optimizer	Adam	cAdam		41.7	58.3		3.458	1.781		cAdam
learning rate	0.01	0.001		63.9	36.1		3.241	11.151		0.01

Table 5.6: parameter influence regarding *final val loss mae avg* for the chemReg Adam dataset

5.3 test loss

parameter name	best values				
'test loss avg' (avg)	0.11298	0.11379	0.11389	0.11409	0.11445
test loss mae avg	0.11298	0.11379	0.11389	0.11409	0.11445
final val loss avg	0.012562	0.013544	0.012455	0.0066663	0.0065895
training time avg	566.85	43.583	44.379	576.57	80.447
neurons per layer	(50, 10)	(40, 20)	(50, 10)	(40, 20)	(30, 30, 10)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	MSE	MSE	log cosh	log cosh
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	100	1000	1000	100	1000
optimizer	cAdam	Adam	Adam	cAdam	cAdam
learning rate	0.001	0.01	0.01	0.001	0.01
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.7: best settings regarding *test loss avg* for the chemReg Adam dataset

parameter name	worst values				
'test loss avg' (avg)	0.18477	0.19365	0.19938	0.20834	0.22996
test loss mae avg	0.18477	0.19365	0.19938	0.20834	0.22996
final val loss avg	0.036515	0.019063	0.020418	0.044762	0.051253
training time avg	17.242	17.453	17.19	17.624	16.914
neurons per layer	(50, 10)	(50, 10)	(40, 20)	(30, 30, 10)	(40, 20)
activation functions	sigmoid	sigmoid	sigmoid	sigmoid	sigmoid
last activation function	linear	linear	linear	linear	linear
loss function	MSE	log cosh	log cosh	MSE	MSE
training data percentage	1	1	1	1	1
number of epochs	500	500	500	500	500
batch size	10000	10000	10000	10000	10000
optimizer	Adam	Adam	Adam	Adam	Adam
learning rate	0.001	0.001	0.001	0.001	0.001
ε	10^{-7}	10^{-7}	10^{-7}	10^{-7}	10^{-7}

Table 5.8: worst settings regarding *test loss avg* for the chemReg Adam dataset

parameter name	parameter values			win ratios in %			avg. differences in %			best value
neurons per layer	(40, 20)	(50, 10)	(30, 30, 10)	27.1	50.0	22.9	3.242	1.677	5.055	(50, 10)
activation functions	ReLU	sigmoid		55.6	44.4		2.313	4.219		ReLU
loss function	MSE	log cosh		37.5	62.5		2.93	1.147		log cosh
batch size	100	1000	10000	50.0	47.9	2.1	4.201	4.151	25.616	*unclear*
optimizer	Adam	cAdam		45.8	54.2		3.551	1.499		cAdam
learning rate	0.01	0.001		65.3	34.7		3.06	10.756		0.01

Table 5.9: parameter influence regarding *test loss avg* for the chemReg Adam dataset