Effect of Sample Size on Kernel Deep Neural Network

Xiaoran Tong
Qin Lu

Michigan State University

tongxia1@msu.edu
qlu@epi.msu.edu

December 15, 2017

Table of Content

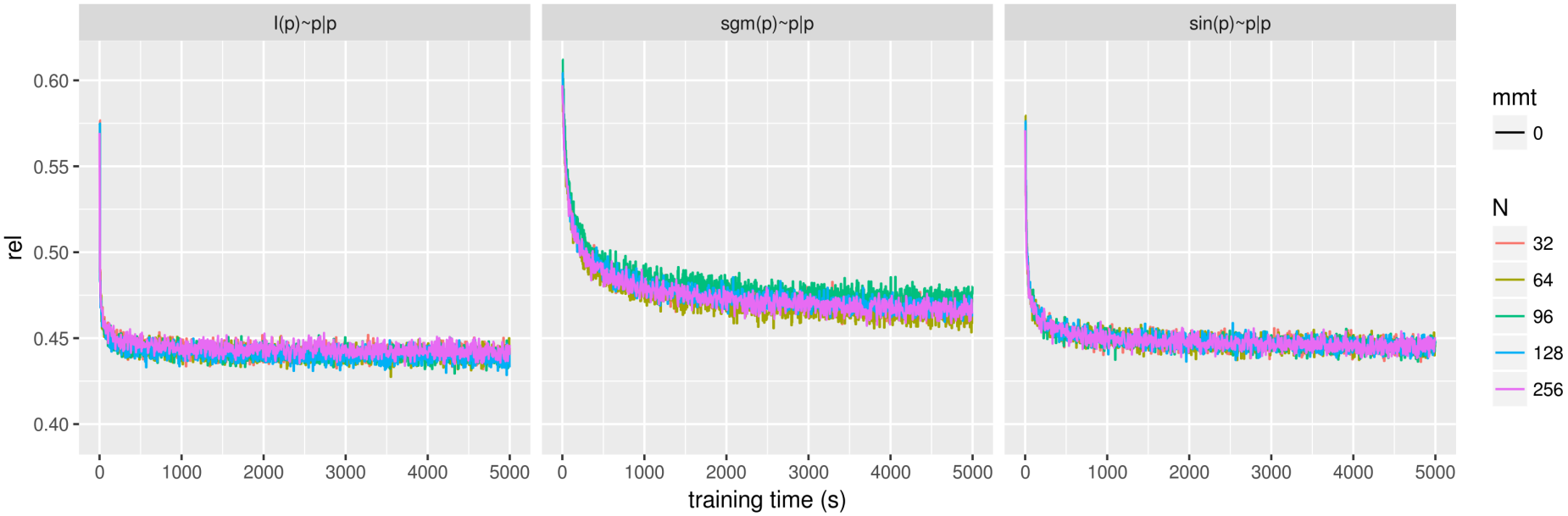
EXP 01: Tied Weight, $h(linear) \sim linear|linear$

Simulation Settings:

- N = {32, 64, 96, 128, 256}
- P=2000, FRQ=0.50, ϕ=2.0
- link (h) = {I, sigmoid, sin}
- Number of Hidden Units = 96
- True Model: $y \sim \mathcal{N}(0, X'X + \phi I)$
- Batch Size = 16
- learning rate = 6.4e-4, Momentum = 0.0
- Reserved Time = 2 hours, Reserved Iteration = ∞
- No rendom seed.

EXP 01: Tied Weight, $h(linear) \sim linear|linear$

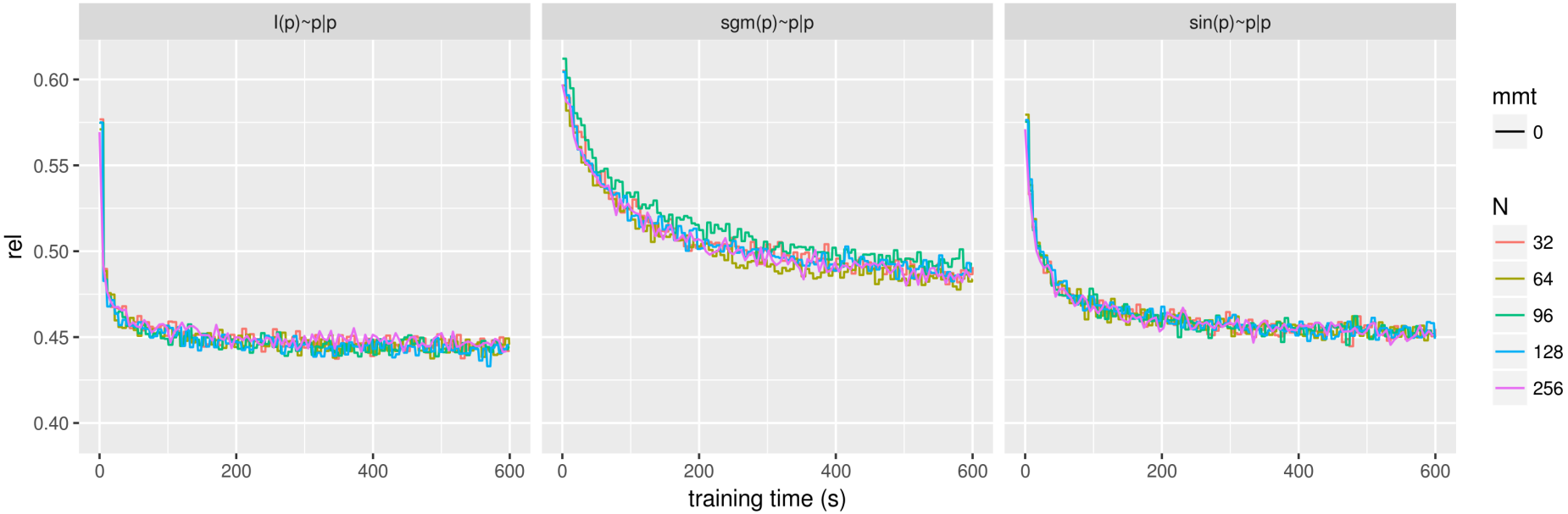
Performance by Time, till finish:



- y-axis: evaluation error (H=256) relative to NULL model.
- no significant performance differenciation between choice of sample sizes.

EXP 01: Tied Weight, $h(linear) \sim linear|linear$

Performance by Time, first 10 min:



- y-axis: evaluation error (H=256) relative to NULL model.
- no significant performance difference among choices of sample size.

EXP 01: Tied Weight, $h(linear) \sim linear|linear$

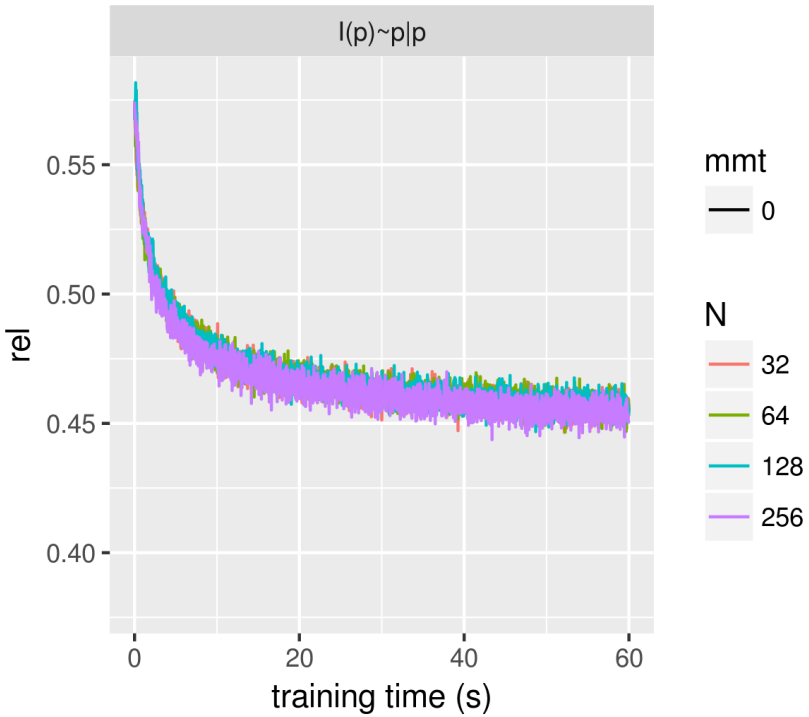
Evaluation error at Finish:

	N	P	H	frq	ycv	lnk	PHI	bsz	M	err	rep
1	32	2000	256	0.50	p	I	2	16	96	0.92	49
2	64	2000	256	0.50	p	I	2	16	96	0.90	49
3	96	2000	256	0.50	p	I	2	16	96	0.91	50
4	128	2000	256	0.50	p	I	2	16	96	0.90	49
5	256	2000	256	0.50	p	I	2	16	96	0.90	48

- no significant performance difference among choices of sample size;
- samaller batch improves stability.

EXP 02: Tied Weight, $h(linear) \sim linear|linear$

Performance by Time, first 1 min, finner recording:



- y-axis: evaluation error (H=256) relative to NULL model.
- no significantly different performce by varying sample sizes.

EXP 01: Tied Weight, $h(linear) \sim linear|linear$

Evaluation error at Finish, finner recording:

	N	P	H	frq	ycv	lnk	PHI	bsz	M	mu	rep
1	32	2000	256	0.50	p	I	2	16	96	0.92	49
2	64	2000	256	0.50	p	I	2	16	96	0.90	49
3	96	2000	256	0.50	p	I	2	16	96	0.91	50
4	128	2000	256	0.50	p	I	2	16	96	0.90	49
5	256	2000	256	0.50	p	I	2	16	96	0.90	48

- no significant performance difference among choice of sample sizes;

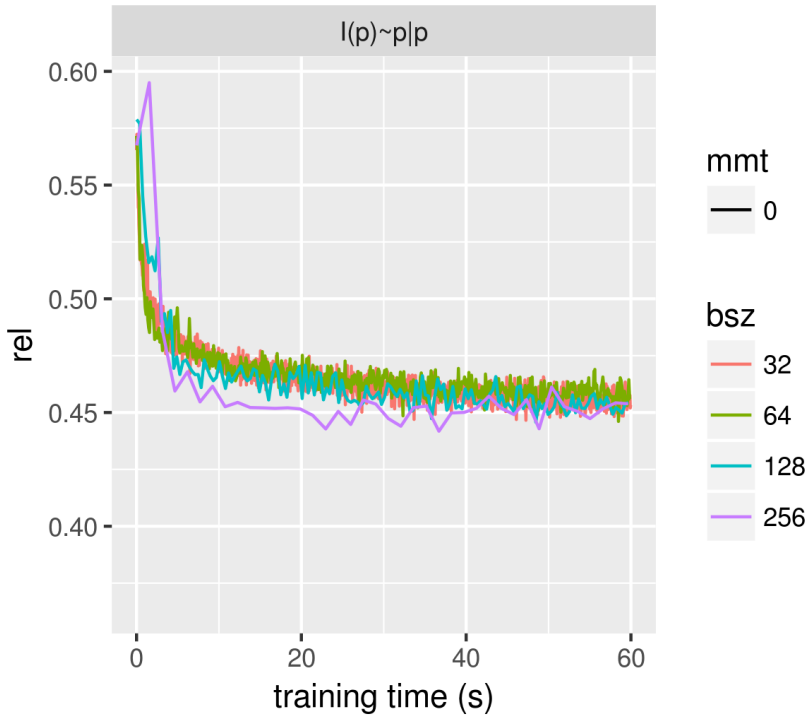
EXP 01: Tied Weight, *hlinear* $\sim$ *linear*|*linear*

Simulation Settings:

- N = 256, H=256
- P=2000, FRQ=0.50, ϕ=2.0
- link (h) = I
- Number of Hidden Units = 96
- True Model: $y \sim \mathcal{N}(0, X'X + \phi I)$
- Batch Size = {32, 64, 128, 256}
- learning rate = {6.4e-4, 3.2e-4, 1.6e-4, 8.0e-5}
- Momentum = 0.0
- Reserved Time = 2 hours, Reserved Iteration = ∞
- No rendom seed.

EXP 01: Tied Weight, *hlinear* $\sim$ *linear*|*linear*

Performance by Time, first 1 min:



- y-axis: evaluation error (H=256) relative to NULL model
- larger batch may converge better at the begining.

EXP 01: Tied Weight, *hlinear* $\sim$ *linear*|*linear*

Performance at Finish:

	N	P	H	frq	ycv	lnk	PHI	bsz	M	mu	rep
1	256	2000	256	0.50	p	l	2	32	96	0.90	46
2	256	2000	256	0.50	p	l	2	64	96	0.90	42
3	256	2000	256	0.50	p	l	2	128	96	0.93	33
4	256	2000	256	0.50	p	l	2	256	96	0.94	31

- large batch is better at the begining (1 min), but worse at finishing (2 hours).
- larger batch is untable, a new simulation is ongoing.

- by varying sample size, the evaluation error after convering seems no difference.