

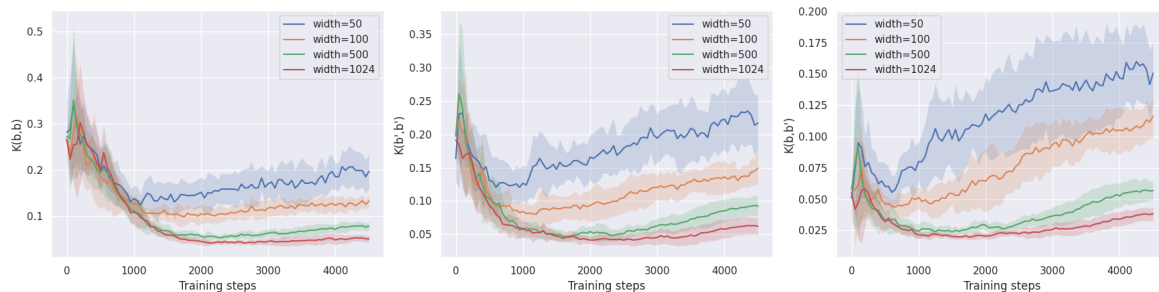
# Online Limited Memory Neural-Linear Bandits with Likelihood Matching - Supplementary Material

Ofir Nabati<sup>1</sup> Tom Zahavy<sup>1,2</sup> Shie Mannor<sup>1,3</sup>

## A. Representation drift experiment

The NTK was sampled during training with the Shuttle Statlog dataset (Newman et al., 2008). We compute the NTK for two fixed contexts  $b$  and  $b'$ , taken from the dataset. Each context is composed of 9 features describing the space shuttle flight. The goal is to predict the state of the radiator of the shuttle (the reward). There are  $N = 7$  possible actions; for correct predictions the reward is  $r = 1$  and  $r = 0$  otherwise.

The weights of the network were initialized according to (Jacot et al., 2018). We ran the experiments for various network widths: 50, 100, 500, and 1024. The network was trained as described in the main paper (Method and setup) with a random policy. The NTK was sampled every 50 iterations during training for the estimated reward of the first action. The graphs below present  $K(b, b)$ ,  $K(b', b')$  and  $K(b, b')$  from left to right accordingly during training averaged over 10 seeds.



## B. Raw results

Name	d	A	Full memory		Limited memory			NTK based		
			LinearTS	NeuralLinear	LiM2 (Ours)	NeuralLinear-MM	NeuralLinear-Naive	NeuralUCB	NeuralTS	NeuralLinear-NTK
Mushroom	117	2	<b>11162 ± 1167</b>	10810 ± 428	9880 ± 1776	4602 ± 1408	4843 ± 1228	-32 ± 3	-34 ± 18	9785 ± 1012
Financial	21	8	3752 ± 8	3560 ± 12	<b>3762 ± 18</b>	2802 ± 21	2726 ± 23	1116 ± 824	876 ± 595	3610 ± 18
Jester	32	8	<b>15944 ± 170</b>	14731 ± 304	14926 ± 571	11940 ± 1307	11647 ± 1066	13397 ± 7	13397 ± 8	14642 ± 200
Adult	88	2	4008 ± 14	4003 ± 12	<b>4043 ± 15</b>	3483 ± 33	3477 ± 16.5	3768 ± 2	3769 ± 2	3990 ± 17
Coverttype	54	7	<b>2961 ± 25</b>	2742 ± 40	2719 ± 62	2241 ± 30	2272 ± 37	1870 ± 11	1877 ± 83	2708 ± 31
Census	377	9	1801 ± 15	2510 ± 21	<b>2827 ± 22</b>	2099 ± 39	2114 ± 34	2019 ± 94	1926 ± 76	2517 ± 42
Statlog	9	7	4460 ± 19	4729 ± 6	<b>4820 ± 68</b>	4545 ± 34	4476 ± 19	4075 ± 3	4348 ± 265	4722 ± 12
Epileptic	178	5	1204 ± 29	<b>1734 ± 46</b>	1501 ± 115	1411 ± 30	1368 ± 26	1010 ± 2	1011 ± 2	1431 ± 47
Smartphones	561	6	3092 ± 18	4208 ± 23	<b>4313 ± 46</b>	2647 ± 17	2626 ± 16	2214 ± 1548	3166 ± 1113	4191 ± 32
Scania Trucks	170	2	4710 ± 171	4837 ± 132	4856 ± 111	4574 ± 220	4650 ± 103	4919 ± 19	<b>4922 ± 23</b>	4730 ± 179
Amazon	7K	5	-	3024 ± 25	<b>3052 ± 160</b>	2793 ± 41	2804 ± 41	-	-	3014 ± 37

Table 1. Cumulative reward on 11 real world datasets.

<sup>1</sup>Department of Electrical-Engineering, Technion Institute of Technology, Israel <sup>2</sup>DeepMind <sup>3</sup>Nvidia Research. Correspondence to: Ofir Nabati <ofirnabati@gmail.com>.

## References

- Jacot, A., Gabriel, F., and Hongler, C. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in neural information processing systems*, pp. 8571–8580, 2018.
- Newman, D., Smyth, P., Welling, M., and Asuncion, A. U. Distributed inference for latent dirichlet allocation. In *Advances in neural information processing systems*, pp. 1081–1088, 2008.