

Supplementary for Learning guarantee of reward modeling using deep neural networks

1 AN EXPERIMENT ON SYNTHETIC DATA

We construct a synthetic experiment to illustrate our theory in the guidance for deep neural network implementation. In this section, we consider two parametrization of the deep reward modeling using Bradley-Terry and Thurstonian models, respectively. We refer to the Example 1 and 2 for more details. This reward function is specified as: $r^*(s, a_1) = 2 \sin(4\phi(s)^\top w^*)$ and $r^*(s, a_0) = -2 \sin(4\phi(s)^\top w^*)$ where $\phi(s) = (\sin(s_1), \dots, \sin(s_d))$ is a non-parametric transformation for creating non-linearity. The identification condition, $r^*(s, a_1) + r^*(s, a_0) = 0$ for every given s , is ensured. Furthermore, as demonstrated in expression (1) and (2), both regret and preference depend solely on the difference between rewards. Accordingly, in our implementation, we configure the output of the neural network to directly represent the reward difference, $\hat{r}(s, a_1) - \hat{r}(s, a_0)$, rather than estimating individual rewards separately.

We generate n state observations s , with each sampled independently from a uniform distribution over $[0, 1]^d$. In this example, we consider each dataset with the dimension of $d = 10$ and sample size $(n_{train}, n_{eval}, n_{test}) = (2^{10}, 2^9, 2^{10})$. The true weight w^* is fixed as $(-0.040, 1.726, -0.814, 1.372, 0.506, -0.482, -0.785, 0.668, -0.443, 0.188)^\top$, which is randomly generated *a priori*. We evaluate the rectangular neural networks where all the hidden layers are of the same width. Specifically, we consider the networks with widths $\{2^i, i = 4, \dots, 12\}$ and the depths ranging from 3 to 13. The number of parameters ranges from about 500 to roughly 2×10^7 . We note that the candidate networks have varying expression powers and are sufficient to validate the guidance ability of our theory in the network design. Each configuration is evaluated across 50 independent replications. The averaged regret results are presented in Figure 1.

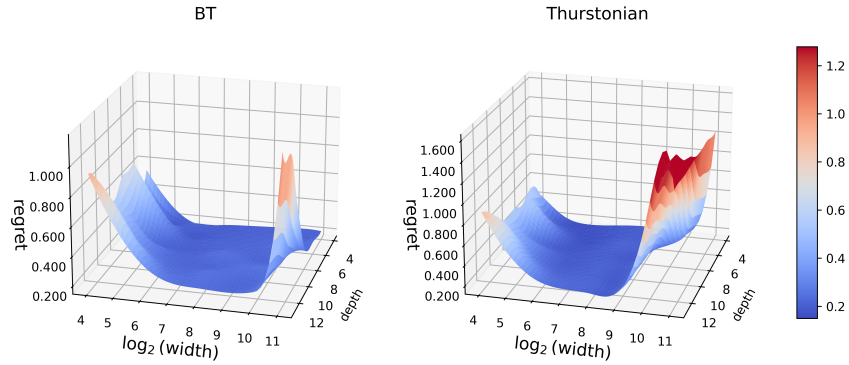


Figure 1: Regrets for synthetic data under different neural network configurations.

Recall that our theoretical analyses reveal a crucial trade-off in neural network architecture design. The interplay between stochastic and approximation errors fundamentally impacts model performance. Proposition 1 and Proposition 2 establish that while increased network complexity reduces approximation error, it simultaneously amplifies stochastic error under finite data scenarios. This finding emphasizes the importance of balanced architecture selection.

Our empirical results (see Figure 1) provide compelling evidence for this theoretical framework. Using a fixed sample size, we examined regret across varying network configurations. We observed a non-monotonic relationship: initially, deeper and wider networks reduce regret as the approximation capability of networks increases. However, beyond the near-optimal network configuration, further increases in model complexity led to degraded performance. This is consistent with our theory that stochastic error dominates under this case; this empirical evidence is in line with Han et al. (2023).

Furthermore, our results reveal that comparable performance levels can be achieved across diverse architectural configurations, highlighting the adaptability of deep neural networks to varying function complexities. This is evidenced by a relatively flat region in the parameter space where the regret remains near-minimal. Such architectural flexibility is particularly valuable in practice, as it suggests that precise knowledge of the smoothness parameter β is not critical for achieving strong empirical performance. This robustness effectively addresses the common practical challenge of architecture selection under unknown function complexity. We recommend Jiao et al. (2023); Lee et al. (2019) for a more comprehensive analysis across different models.

REFERENCES

- Jinhui Han, Ming Hu, and Guohao Shen. Deep neural newsvendor. [arXiv preprint arXiv:2309.13830](#), 2023.
- Yuling Jiao, Guohao Shen, Yuanyuan Lin, and Jian Huang. Deep nonparametric regression on approximate manifolds: Nonasymptotic error bounds with polynomial prefactors. [The Annals of Statistics](#), 51(2):691–716, 2023.
- Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. [Advances in neural information processing systems](#), 32, 2019.