# Multitask Learning for Neural Network Regularization

Julia Nakhleh
Advisor: Robert Nowak
University of Wisconsin-Madison
jnakhleh@wisc.edu

## Abstract:

Multi-task learning [1], a paradigm in which machine learning models are trained to perform multiple tasks simultaneously, has been shown empirically to improve model generalization performance on predictive tasks in many domains [2]. However, rigorous theoretical understanding of the benefits of multitask learning for neural networks is still largely lacking. Our recent work focuses on a particular instance of multi-task learning: single-layer, finite-width ReLU neural networks with multiple outputs, each of which is trained to fit a different set of labels for the same dataset. We view the first set of labels as the "true" labels (obtained from some real-world dataset), and the remaining sets of labels as artificial auxiliary labels which can be numerically generated according to specification. We show that, when the data is univariate and the auxiliary labels meet mild geometric assumptions, training such a network to interpolate the data points with minimal vector-valued variation (VV) norm [3] will result in the first network output—which fits the "true" dataset—learning the connect-the-dots interpolant of the dataset. We also demonstrate experimentally that for multivariate data, training using the same procedure with Gaussian auxiliary labels leads the first output to learn functions which are on average smoother (in that they have a smaller average gradient norm) than those obtained using traditional weight decay regularization on the first output alone. This analysis of multi-task learning in the relatively simple case of single-layer neural networks and low-dimensional data may lend insight into the theoretical guarantees provided by multi-task learning, and how this paradigm compares to other neural network regularization strategies such as weight decay.

### References

[1] R. Caruana, "Multitask Learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997, doi: 10.1023/A:1007379606734.

[2] Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022, doi: 10.1109/TKDE.2021.3070203.

[3] J. Shenouda, R. Parhi, K. Lee, and R. D. Nowak, "Vector-Valued Variation Spaces and Width Bounds for DNNs: Insights on Weight Decay Regularization." arXiv, May 25, 2023. doi: 10.48550/arXiv.2305.16534.