DATA PARALLELISM: HOW TO TRAIN DEEP LEARNING MODELS ON MULTIPLE GPUS

LAB 1, PART 2: MORE REALISTIC NETWORKS

NVIDIA

DEEP LEARNING INSTITUTE

MODERN NEURAL NETWORKS

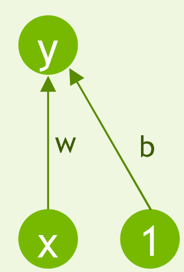
How do they differ from our trivial example?

Not significantly!

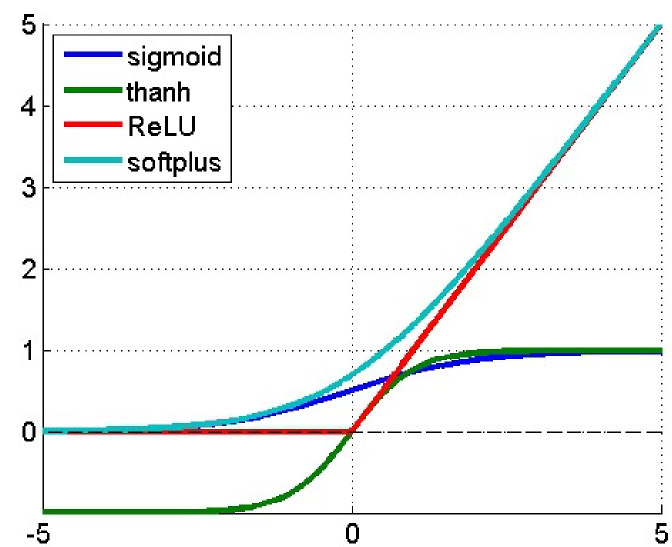
MODERN NEURAL NETWORKS

How do they differ from our trivial example?

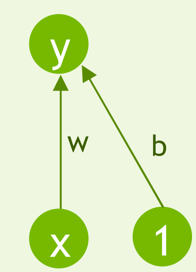
$y = w \cdot x + b$



Nonlinearity



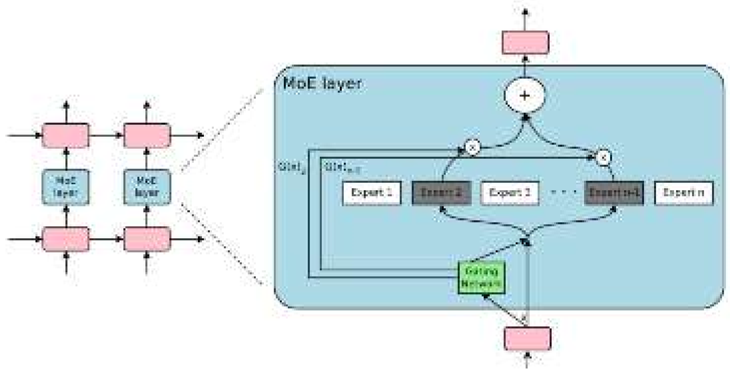
$y = f(w \cdot x + b)$

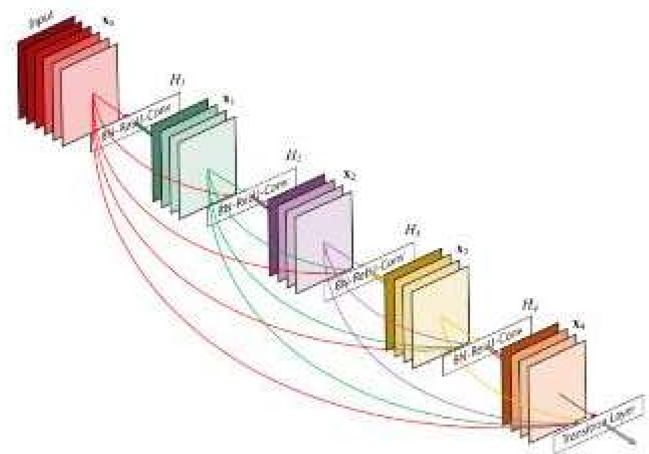


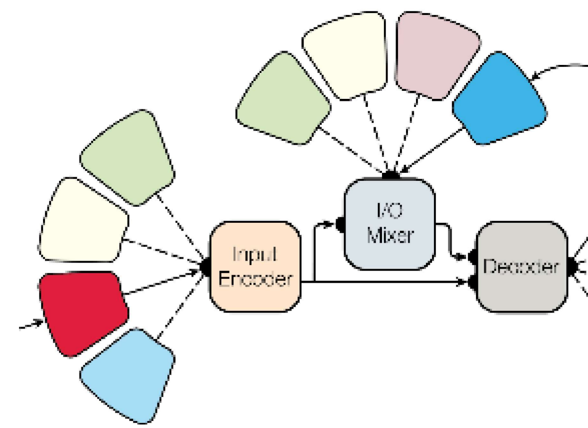
MODERN NEURAL NETWORKS

How do they differ from our trivial example?

More complex interconnection and many more parameters



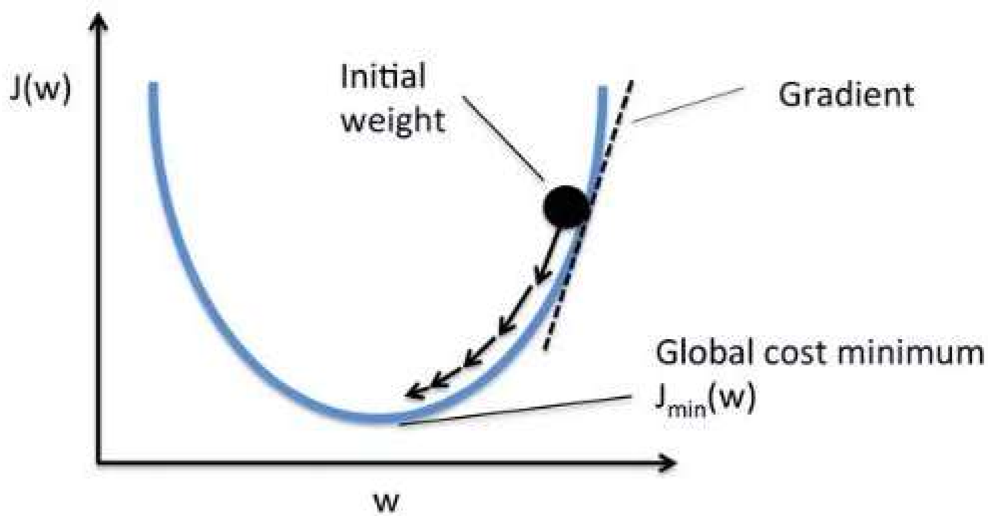


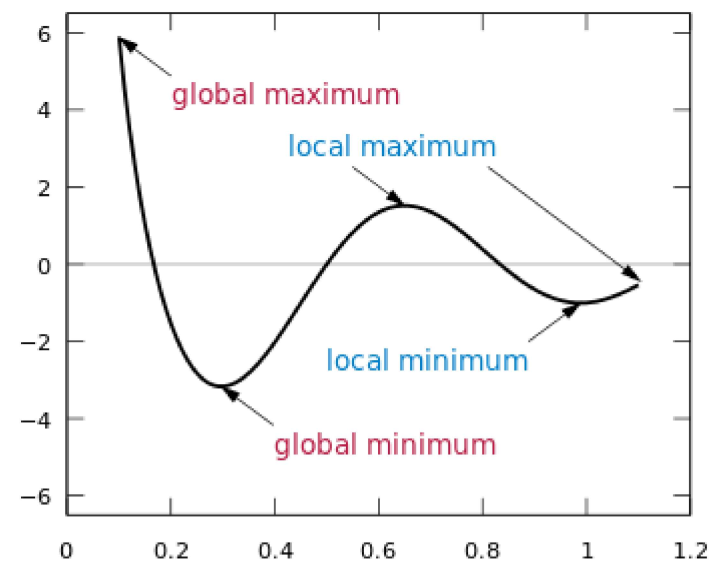


Kaiser, L., Gomez, A. N., Shazeer, N., Vaswani, A., Parmar, N., Jones, L., & Uszkoreit, J. (2017). One model to learn them all. *arXiv preprint arXiv:1706.05137*.
Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869.
Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.0*

NON-CONVEX LOSS FUNCTIONS

Those differences make the optimization problem much more difficu

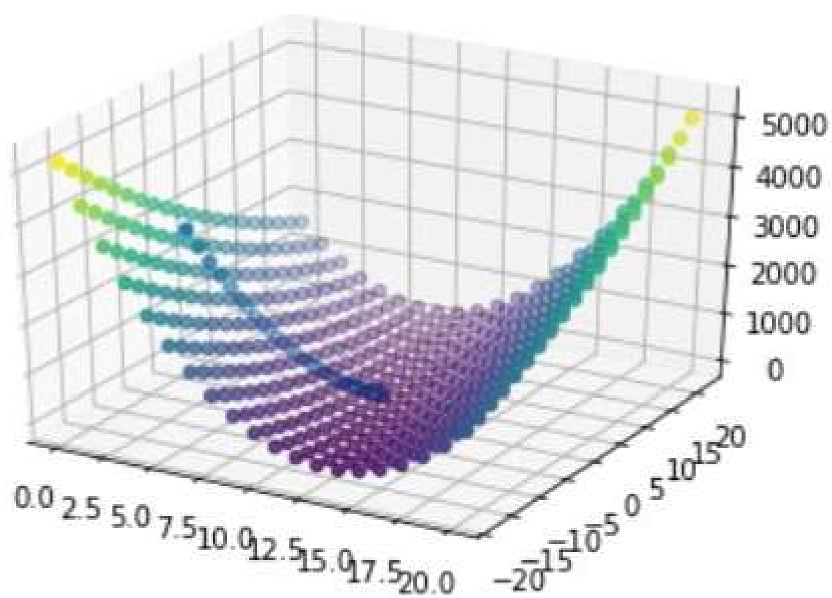




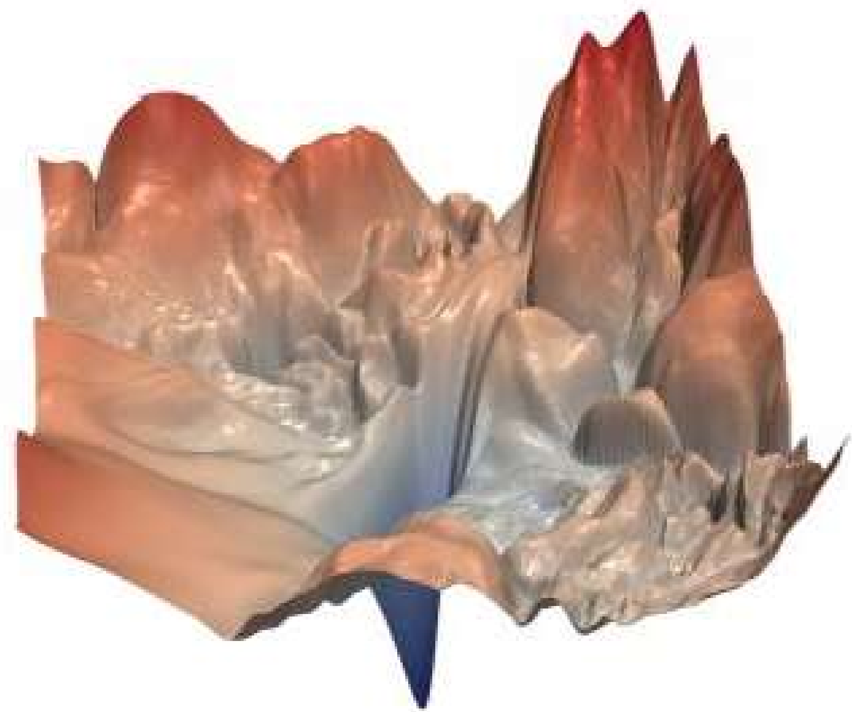
NON-CONVEX LOSS FUNCTIONS

Those differences make the optimization problem much more difficu

Linear model loss function



ResNet-56 loss function projection to 3D – no sk
connections

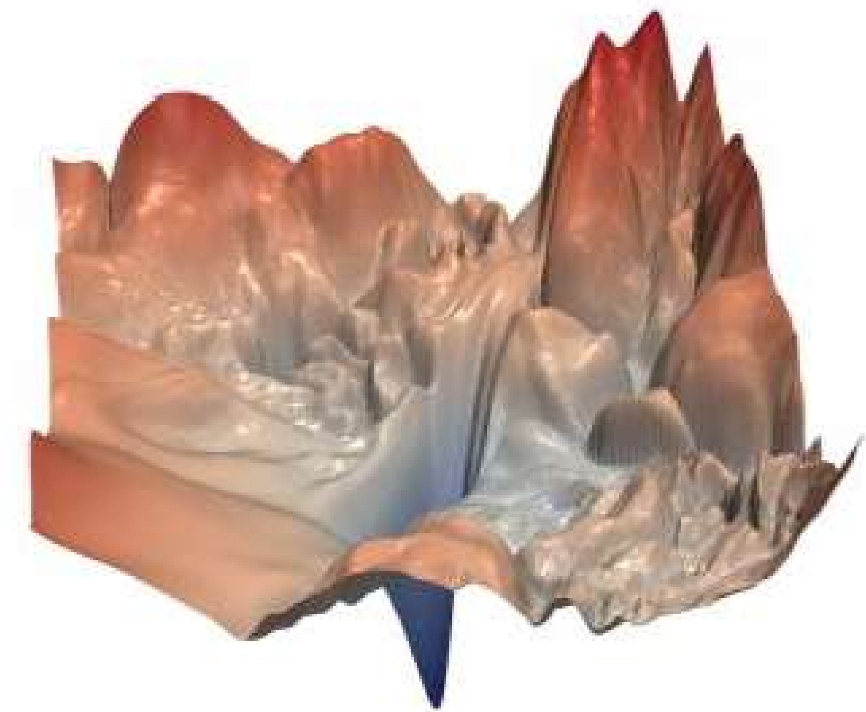


Li, H., Xu, Z., Taylor, G., & Goldstein, T. (2017). Visualizing the Loss Landscape of Neural Nets. arXiv:1712.09913.

NON-CONVEX LOSS FUNCTIONS

Those differences make the optimization problem much more difficu

ResNet-56 loss function projection to 3D – no sk
connections

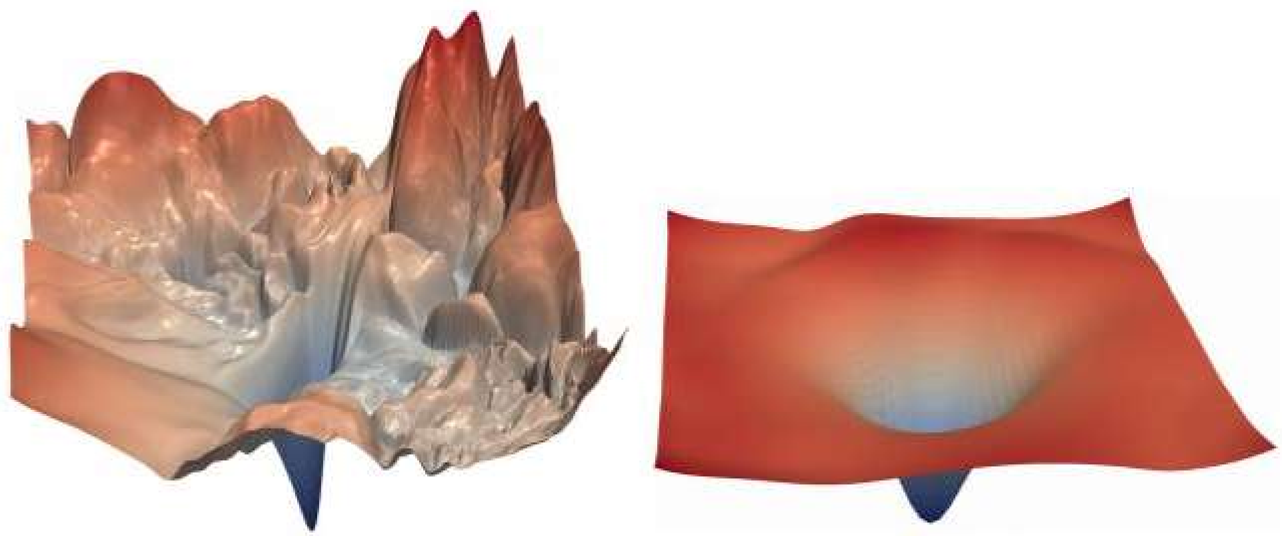


Why do we succeed in finding good local minima?

Li, H., Xu, Z., Taylor, G., & Goldstein, T. (2017). Visualizing the Loss Landscape of Neural Nets. *arXiv:1712.09913*.

NON-CONVEX LOSS FUNCTIONS

Recent advances such as residual connections simplify optimization



(a) without skip connections (b) with skip connections

Li, H., Xu, Z., Taylor, G., & Goldstein, T. (2017). Visualizing the Loss Landscape of Neural Nets. arXiv:1712.09913.