

# FUNDAMENTALS OF DEEP LEARNING FOR MULTI-GPUS

LAB 1, PART 2: MORE REALISTIC NETWORKS



DEEP  
LEARNING  
INSTITUTE

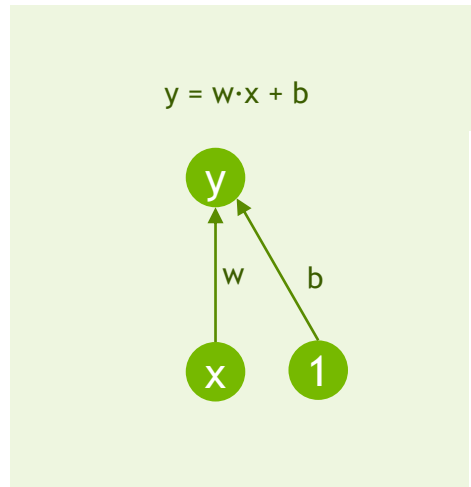
# MODERN NEURAL NETWORKS

How do they differ from our trivial example?

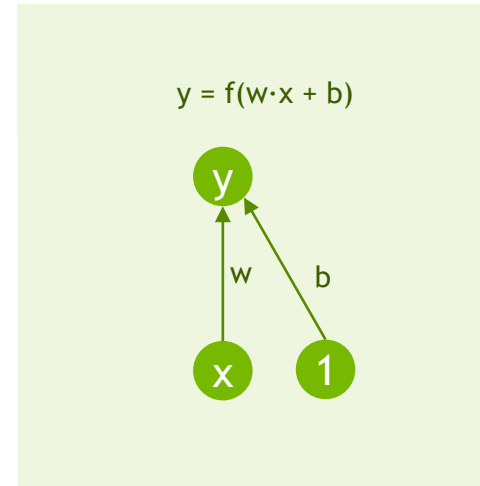
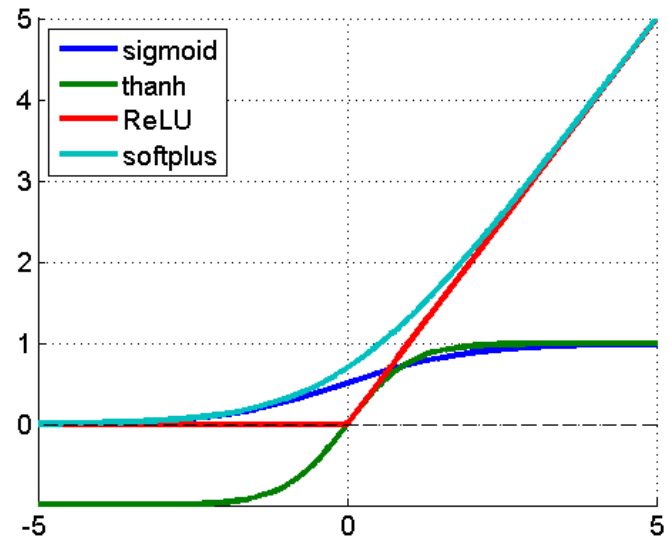
Not significantly!

# MODERN NEURAL NETWORKS

How do they differ from our trivial example?



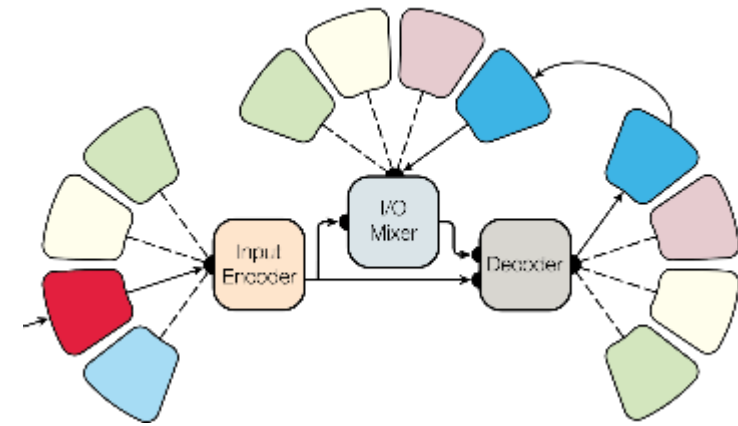
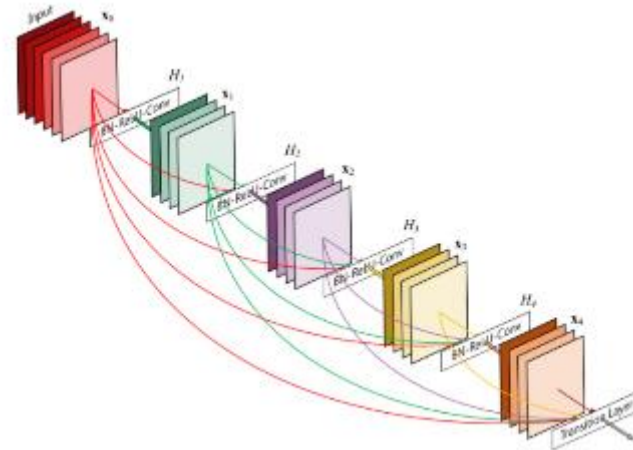
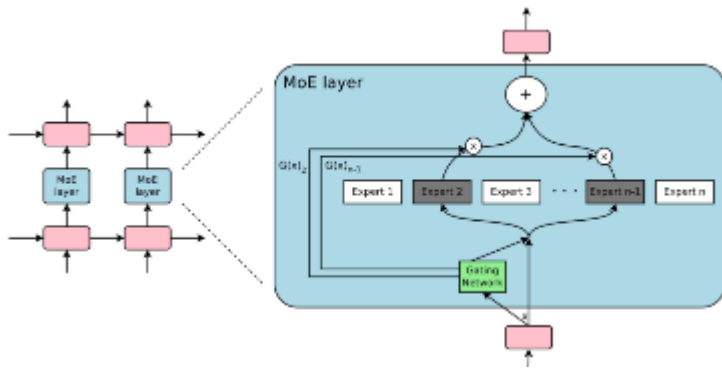
Nonlinearity



# MODERN NEURAL NETWORKS

How do they differ from our trivial example?

More complex interconnection and many more parameters



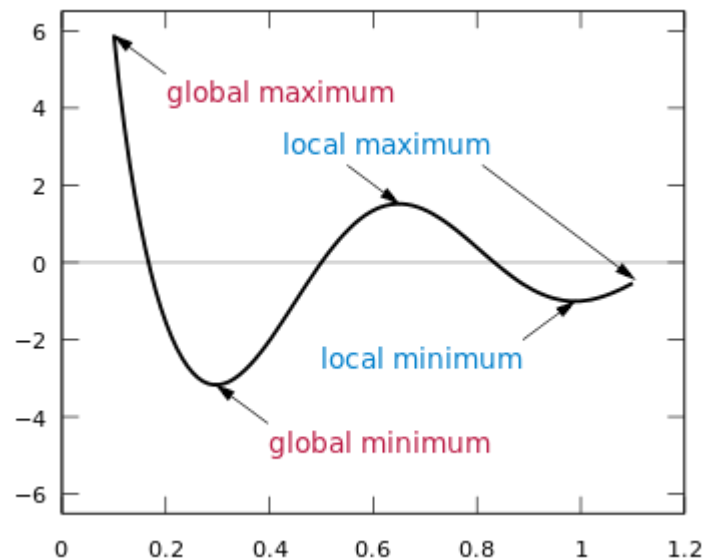
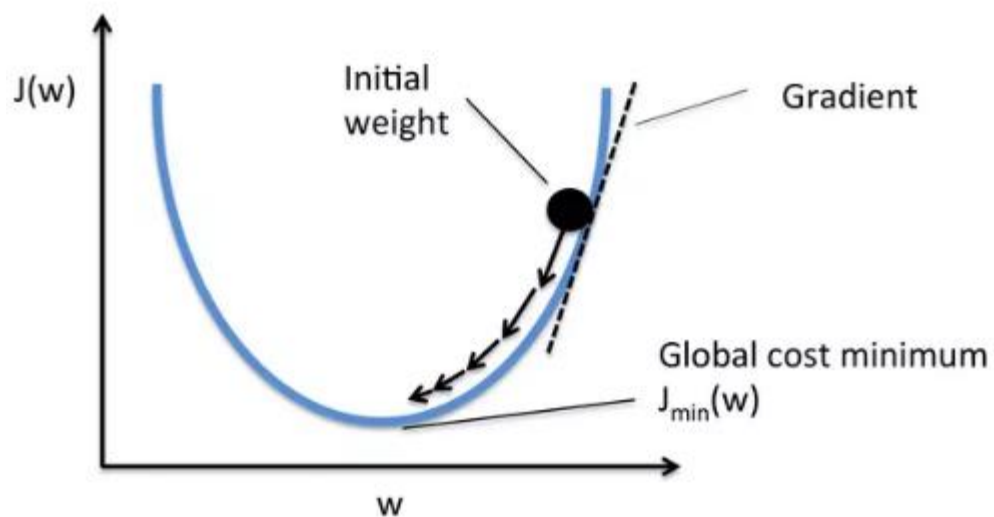
Kaiser, L., Gomez, A. N., Shazeer, N., Vaswani, A., Parmar, N., Jones, L., & Uszkoreit, J. (2017). One model to learn them all. *arXiv preprint arXiv:1706.05137*.

Andolina, 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:1701.06538*.

# NON-CONVEX LOSS FUNCTIONS

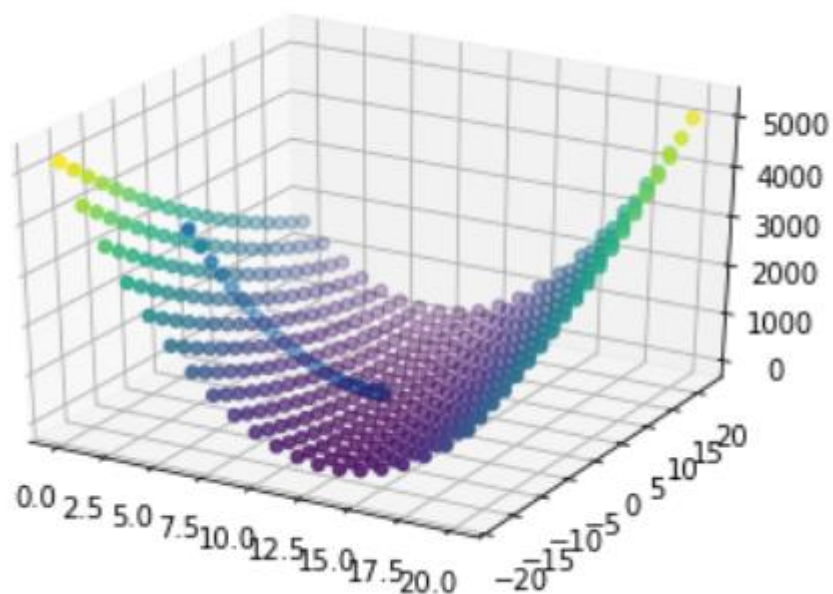
Those differences make the optimization problem much more difficult



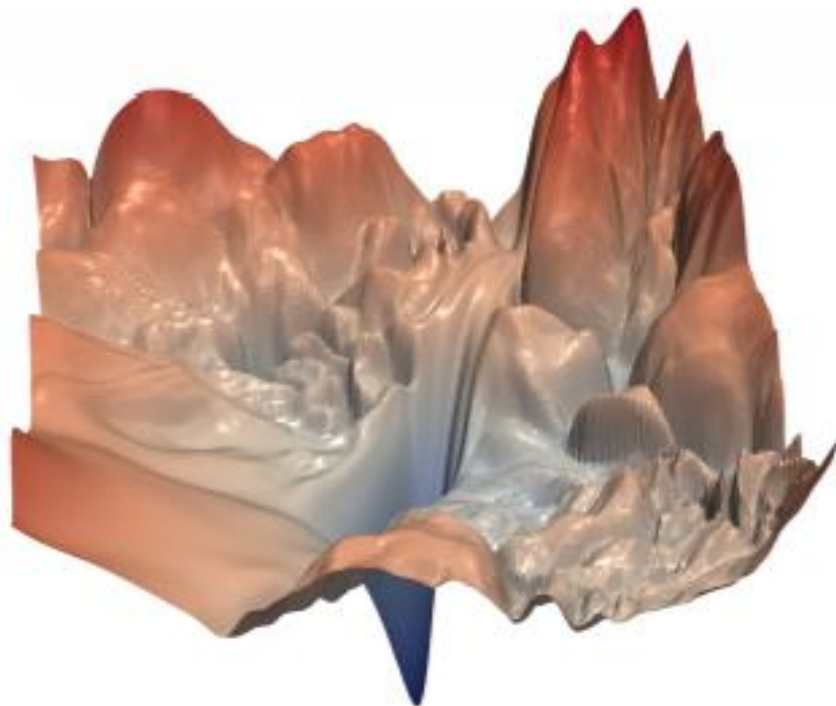
# NON-CONVEX LOSS FUNCTIONS

Those differences make the optimization problem much more difficult

Linear model loss function



ResNet-56 loss function projection to 3D - no skip connections

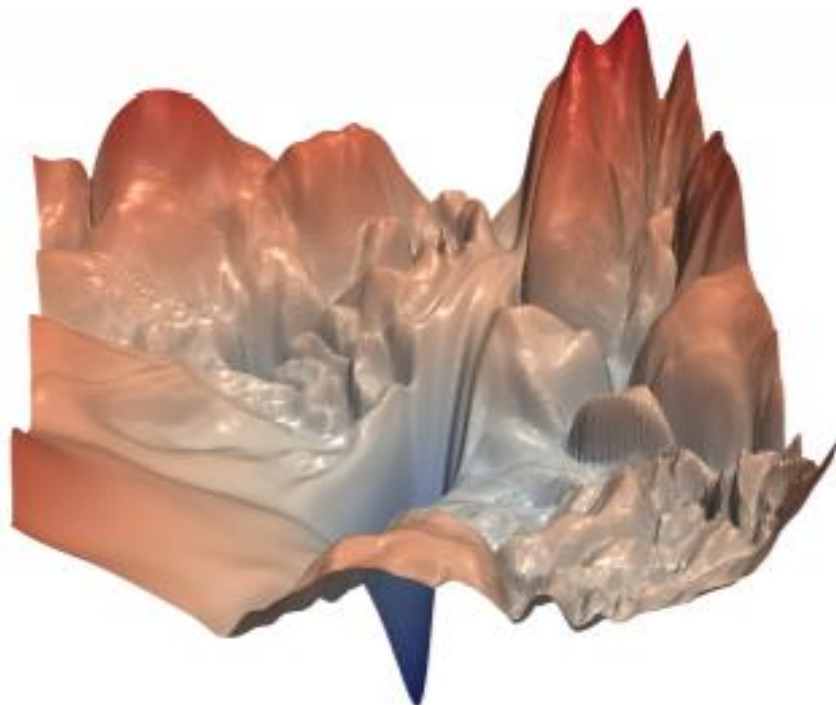


# NON-CONVEX LOSS FUNCTIONS

Those differences make the optimization problem much more difficult

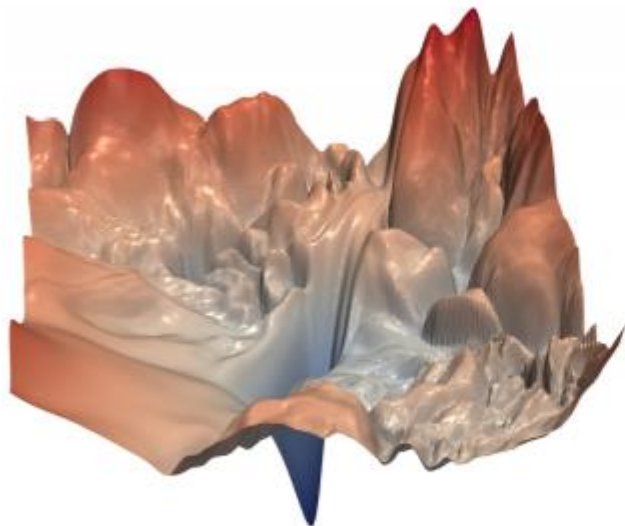
ResNet-56 loss function projection to 3D - no skip connections

Why do we succeed in finding good local minima?

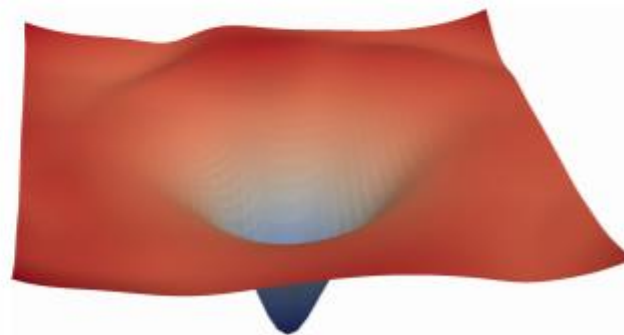


# NON-CONVEX LOSS FUNCTIONS

Recent advances such as residual connections simplify optimization

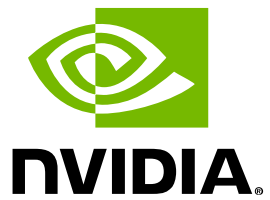


(a) without skip connections



(b) with skip connections





DEEP  
LEARNING  
INSTITUTE

[www.nvidia.com/dli](http://www.nvidia.com/dli)