Sparse Neural Networks: Improving Feedforward Neural Network Efficiency with Sparse Matrix Constructions

Dr, Choromanski EMAIL@YOURDOMAIN.EDU
Google Research, NY

Alfred Bouérly EMAIL@COAUTHORDOMAIN.EDU
Columbia University

Patrick Boueri EMAIL@COAUTHORDOMAIN.EDU
Columbia University

Abstract

We show empirical evidence that sparse connectivity in between layers in a Feed Forward Neural Network does not impact accuracy significantly, as compared to a fully connected layer. Using the canonical MNIST data set, we compute accuracy measures for many Feed Forward Neural Nets with different connection schemes and topologies, showing there is no significant drop off as low as 10

1. Introduction

- Feed Forward Networks, construction, theory and successes (LeCun et al., 1998)
- Problems with Neural Networks, overspecification?

2. Experimental Setup

2.1. Sparse Construction

- Intro to sparsification (What we mean by Sparse)
- How we impose sparsification in training
- Classes of graphs:
  - Random
  - Circulant
  - Pseudo-Random TBD
- Big O Notation of sparse vs. FC for all

Proceedings of the 32^{nd} *International Conference on Machine Learning*, Lille, France, 2015. JMLR: W&CP volume 37.

2.2. methods

Keraspatal, training params, and hardware

3. Results

Here we discuss the actual measured results in terms of accuracy, and come up with really good metrics of accuracy and a full picture of what the differences are between the Fully Connected and Sparse Paridigm

3.1. Example Image

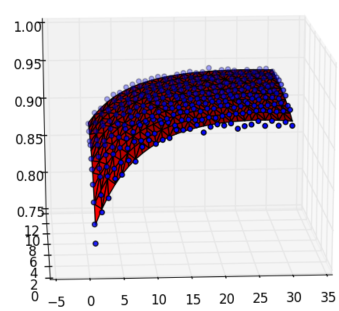

Figure 1. Example Result from Structured.

4. Analysis

Fully Connected Networks do not add as much accuracy, over parameterized, with quick wins coming with very little connections, and additional ones, adding very little. Something like that.

5. Future Work

- Implementing in different Neural Nets (Convolutional neural nets, are in a sense sparse)
- Training analysis (does it take longer, shorter, more or less local minima?)
- Datasets where it succeeds, where it fails
- More information on topologies vs. sparsity
- Computational resources were a big limitation
- Maybe sparsifying intentionally during training (Drop out and Stay out)

6. Conclusion

Bring home major point : most of NN accuracy, comes with few connections
Link to IoT if possible (it's 'hot')

References

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, Pat (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.