Analysis of Non-Gradient Learning in Neural Networks

Dhrumil Mayur Mehta (dm9076@rit.edu) **Advisor(s)**: Prof. Ifeoma Nwogu, Timothy Zee

Computer Science @ RIT
B.Thomas Golisano College of Computing & Information Sciences

Introduction

Disadvantages of backpropagation:

- Two ways mechanism for training neural networks.
- Weight transport problem.
- Exploding/vanishing gradient problems.
- Extensive hyperparameter tuning.

Solution: Train neural networks in a backpropagation-free way.

Background

Target propagation – Local targets for every layers using approximate inverse function [3].

Random feedback alignment – Multiplying feedback error signals with a fixed random matrix instead of weights [1].

Direct feedback alignment – Propagates feedback signal to all the layers directly [2].

Local search – Performs random search optimization on a subset of all dimensions at a time instead of optimizing globally [4].

Methodology

Local Representation Alignment – Local targets for every layers using gradient of next layers' loss with respect to its pre-activation [5].

The computation of target y^{l-1} for layer l - 1 is illustrated as:

$$y^{l-1} = f^{l-1}(h^{l-1})$$

$$h^{l-1} = h^{l-1} - \beta.E_l\,(\partial L^l / \partial h^l)$$

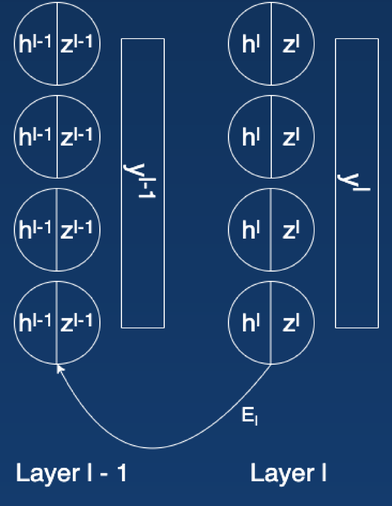

Figure 1 - Layer-wise target update in LRA

Results

Table 1, 2 - Comparative analysis of a convolutional neural net trained on MNIST and Fashion MNIST using backpropagation & LRA.

Dataset	Type	Backpropagation	LRA
MNIST	Training	99.77%	89.13%
	Testing	97.49%	89.45%
Time to execute		11.6 h	36.4 h

Dataset	Type	Backpropagation	LRA
FMNIST	Training	100%	100%
	Testing	88.60%	82.39%
Time to execute		119.9 h	14.6 h

Observations:

- Loss curves in backpropagation decrease more progressively than LRA.
- Learning rate tuning does not affect LRA training.
- LRA can handle poor weight initialization schemes.

Conclusion

- Many alternative methods can be used to avoid drawbacks of backpropagation.
- Local representation alignment is more biologically plausible and avoids exploding/vanishing gradient problems.
- This optimization method can be extended for more complex architectures like residual networks.

References

[1] T. Lillicrap, D. Cownden, D. Tweed, and C. Akerman, "Random synaptic feedback weights support error backpropagation for deep learning, "Nature Communications, vol. 7, no. 13276, pp. 1–10, 2016.

[2] A. Nøkland, "Direct feedback alignment provides learning in deep neural networks," 2016.

[3] D.-H. Lee, S. Zhang, A. Fischer, and Y. Bengio, "Difference target propagation," 2015.

[4] A. Aly, G. Guadagni, and J. B. Dugan, "Derivative-free optimization of neural networks using local search," in2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference(UEMCON), 2019, pp. 0293–0299.

[5] A. G. Ororbia, A. Mali, D. Kifer, and C. L. Giles, "Conducting credit assignment by aligning local representations," 2018.