Improved sign based learning algorithm derived by composite nonlinear Jacobi process

BITS Pilani
Hyderabad Campus

SHRI LAKSHMI.A
M.E, Embedded Systems,
Electrical and Electronics Department

Synopsis

- Project goal
- Learning Scheme of GJRprop
- Simulation Results
- Conclusion

innovate achieve lead

BITS Pilani
Hyderabad Campus

PROJECT GOAL

Challenges of Neural Network

1. Neural network is the reason of intelligence in applications around us.
2. To owe to its superiority, nonlinear network has to be TRAINED intensively.
3. One of the most met challenges that can lowkey its performance is TRAINING SPEED.

Objective

1. Reduce the training speed by improving the convergence rate.
2. Make the network convergence-proof from any initial starting point.

Deliverables

1. Model a novel learning scheme called GJRprop meeting the objectives.
2. Validate its credibility with conventional algorithm categorised as optimal.

innovate achieve lead

BITS Pilani
Hyderabad Campus

LEARNING SCHEME of GJRProp

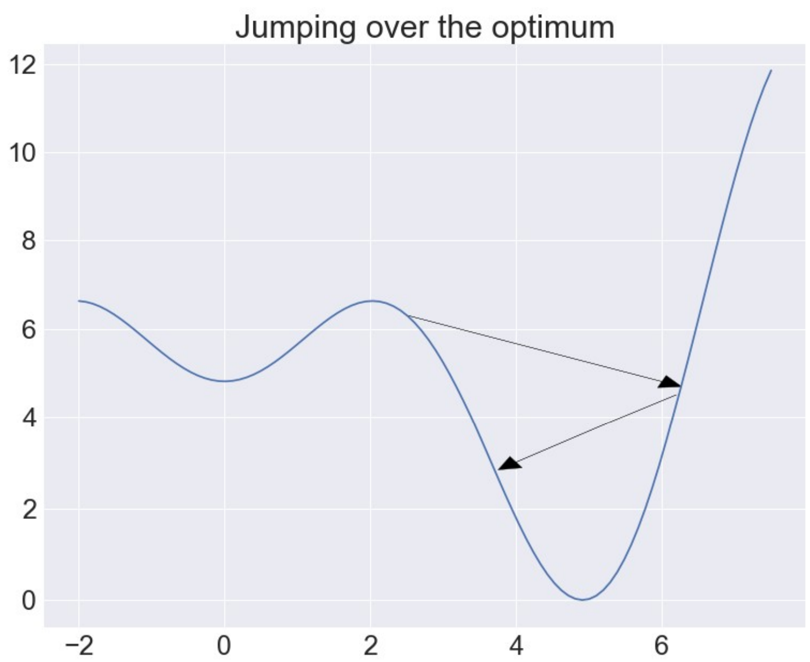
Methodology

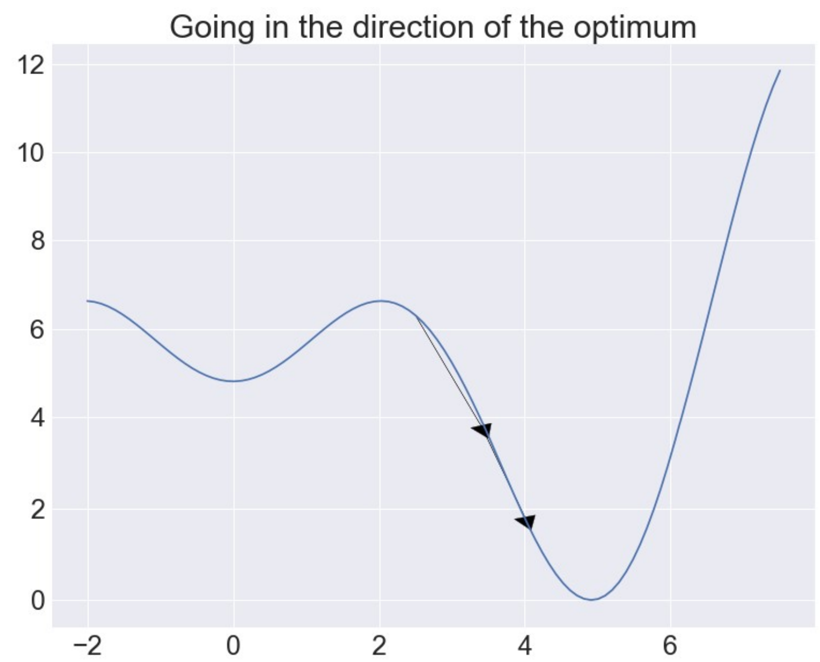
1. Training method for linear network is relatively simpler than nonlinear network.
2. To embrace simplicity in nonlinear case, iterative linear methodology is employed through unconstrained minimization of nonlinear network.
3. Unconstrained minimization: the minima is searched only at the stationary points

Parameter of Focus

1. Dynamic step-size is used to increase the convergence rate.
2. At every iteration, when search direction is desirable, step-size is made higher
3. At undesirable or overleap in search direction, step-size is made smaller.

Idea of dynamic step-size





Ref: [4]

Rprop

Learning heuristic:

1. Dynamic step-size is used in metric update technique.
2. The sign change of the gradient is used as condition check for varying the step-size.

Limitations:

1. The weight update due to sign change do not guarantee error decrement.

Jacobi Bisection algorithm

1. For localization of minima in a interval, interior information is required.
2. Bisection interval halving theorem is used for parameter update technique.

$$x_k^{(t+1)} = x_k^t - h_k\left(\frac{sign\left(\partial_k E(x_k^t)\right)}{2^{(t+1)}}\right); \quad t = 0,1,2 \ldots$$

3. The number of iterations for convergence can be preprocessed from

$$f = \left[log_2[(b_k - a_k)\varepsilon^{-1}]\right]$$

Globally convergent JRprop

Learning methodology:

1. Inculcates dynamic step-size from Rprop.
2. Minima localisation from Jacobian bisection algorithm.
3. Magnitude of error comparison between two consecutive iterations act as condition check.
4. Update technique is as follows:

$$x_{ij}^{(t+1)} = x_{ij}^{t} - \tau^{t}\, diag\left(\eta_1^t, \eta_{2,}^t \dots \eta_{i,}^t, \dots . \eta_{n,}^t\right) sign\left(\nabla E\left(x^t\right)\right)$$

Here, τ^t is the relaxation factor determined by Wolfe's line search algorithm.

Globally convergent GJRprop

Global convergent tip:

The learning rate of one of the elements is modified by

$$\eta_i^t = -\frac{\sum_{\substack{j=1 \\ i \neq j}}^{n} \eta_j^t \, \partial_j E(x^t) + \delta}{\partial_i E(x^t)} \qquad \partial_i E(x^t) \neq 0.$$

1. This convergent tip facilitates minima searching in interval [a,b] set by single step-size update (a) and sum of step- size factored by error magnitude (b)
2. This orients the behavior of both step-size and magnitude of error in determining the search direction.

innovate achieve lead

BITS Pilani
Hyderabad Campus

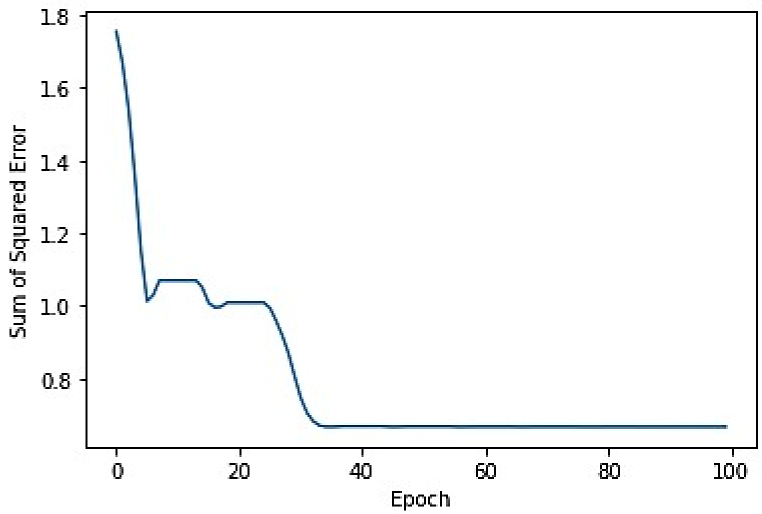
SIMULATION RESULTS

Simulation results

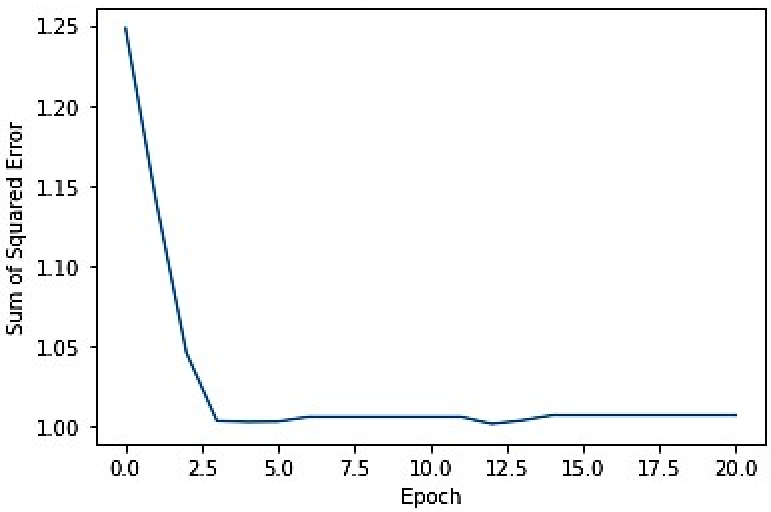
1. Rprop, Jrprop and GJRprop are applied for XOR problem set.
2. The performance metric taken here is Sum of Squared Error.
3. The performance plot is used to study the convergence of the algorithms
4. Convergence accuracy is calculated out of 50 iterations from different starting point.

Performance plot

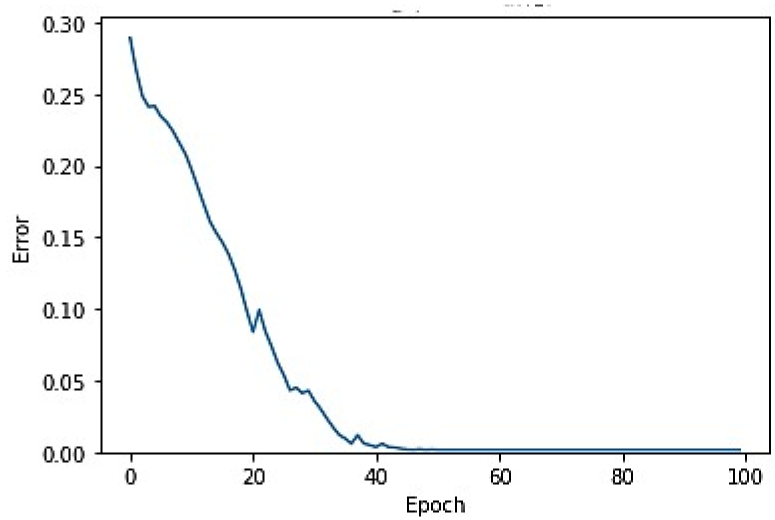
JRprop



GJRprop



Rprop



Comparison table

Algorithm	Rprop	JRprop	GJRprop
Convergence range	35-40	25-35	1.5-2.5
Convergence accuracy	70%	72%	80%

Conclusion

1. From the experimental results, GJRprop shows better results in convergence as compared to other algorithms.
2. The global convergent tip can be used in any algorithm to see improvement in results convergence as a futuristic goal.
3. However, further research is required to determine the performance of the algorithm to determine its possible limitations .

Applications

1. This algorithm can be used for diagnostic tests in medical field: cancer, thyroid problem, E.coli infections
2. The parity check in data communication can be done with the convergence property of GJRprop.
3. It serves mathematical field to determine strong minima and saddle point in problem sets like XOR and similar nonlinear functions.

Reference:

1. Magoulas, G. D., Vrahatis, M. N., & Androulakis, G. S. (1997). Effective backpropagation training with variable stepsize. In *Neural Networks* (Vol. 10, Issue 1, pp. 69–82).
2. Piccinini, G., & Bahar, S. (2013). Neural Computation and the Computational Theory of Cognition. *Cognitive Science*, *37*(3), 453–488.
3. Prechelt, L. (2010). A Set of Neural Network Benchmark Problems and Benchmarking Rules. *Tech. Rep., Universit¨at Karlsruhe, Karlsruhe, Germany*.
4. Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The Rprop algorithm. In {IEEE} International Conference On Neural Networks. *IEEE International Conference on Neural Networks*, *16*(3), 586--591.
5. Robitaille, B., Marcos, B., Veillette, M., & Payre, G. (1996). Modified Quasi-Newton methods for training neural networks. *Computers and Chemical Engineering*, *20*(9), 1133–1140.
6. Thrun, S. (1995). France. *Economic Outlook*, *19*(4), 46–47.
7. Vrahatis, M. N., Magoulas, G. D., & Plagianakos, V. P. (2003). From linear to nonlinear iterative methods. *Applied Numerical Mathematics*, *45*(1), 59–77.
8. Wolfe, P. (1969). *tYn)] t*. *11*(2), 226–235.
9. Zhang, H., Zhang, P., & Hsieh, C. J. (2018). RecurJac: An efficient recursive algorithm for bounding jacobian matrix of neural networks and its applications. *ArXiv*, *Figure 1*.

Reference:

10. Anastasiadis, A. D., Magoulas, G. D., & Vrahatis, M. N. (2005). Sign-based learning schemes for pattern classification. *Pattern Recognition Letters*, *26*(12), 1926–1936.
11. Anastasiadis, A.D., Magoulas, G. D., & Vrahatis, M. N. (2019). A Globally Convergent Jacobi-Bisection Method for Neural Network Training. *International Conference of Computational Methods in Sciences and Engineering 2004 (ICCMSE 2004)*, *April*, 843–848.
12. Anastasiadis, Aristoklis D, Magoulas, G. D., & Vrahatis, M. N. (2003). An efficient improvement of the Rprop algorithm. *Proceedings of the First International Workshop on Artificial Neural Networks in Pattern Recognition (IAPR 2003), University of Florence, Italy*, *June 2014*, 197.
13. Igel, C., & Hüsken, M. (2003). Empirical evaluation of the improved Rprop learning algorithms. *Neurocomputing*, *50*(September 2002), 105–123.
14. Leyffer, S., & Mahajan, A. (2011). Nonlinear constrained optimization: methods and software. *Wiley Encyclopedia of Operations Resaerch and Managment Science*, *February*.
15. Livieris, I. E. (2019). Forecasting economy-related data utilizing weight-constrained recurrent neural networks. *Algorithms*, *12*(4).

THANK YOU.