Developing and Analyzing Neural Network Boosting Methods

Alexander Clinton[1], Dr. Joseph Anderson[2]
[1]University of Notre Dame, [2]Salisbury University

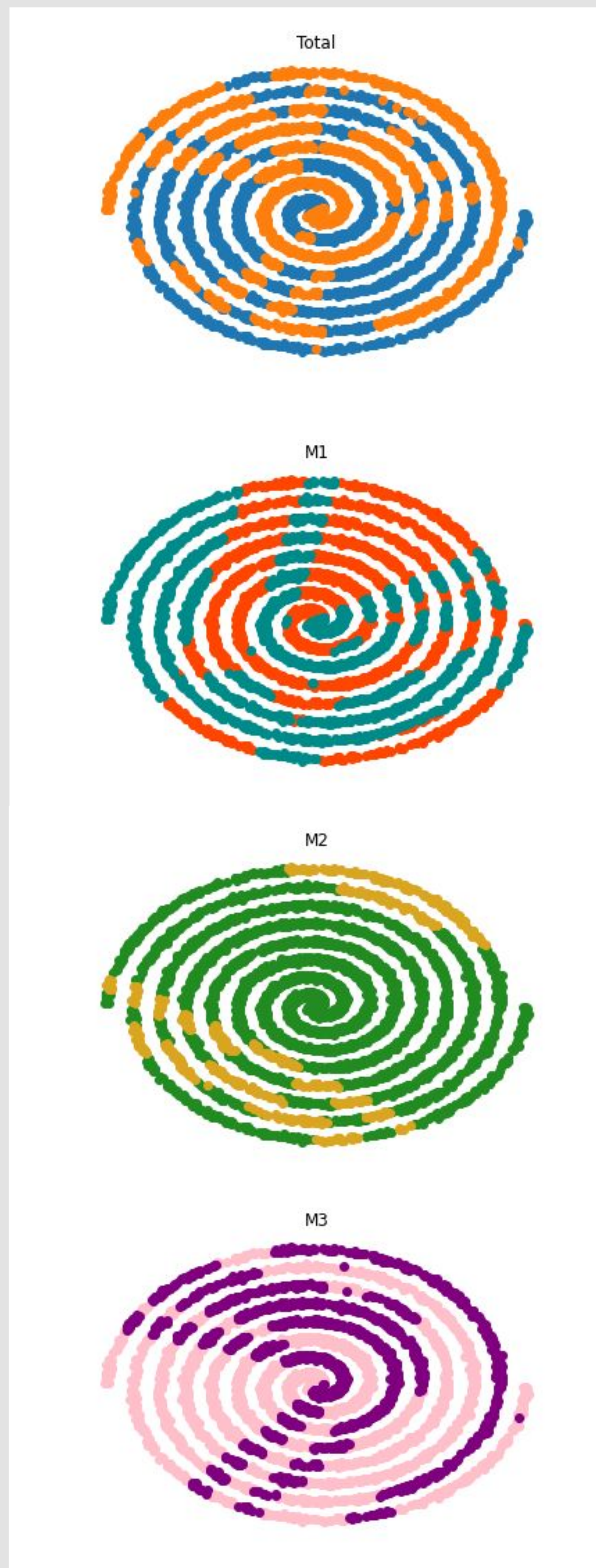
Abstract

Boosting is a term for a series of generic techniques aimed at improving classification problems by using an ensemble of learners. Although the technique is classically used with weak learners such a decision trees, boosting has been effectively applied to advanced predictors such as neural networks. A natural question to ask is whether the benefits of boosting are retained as the learners are able to form more robust hypotheses and mimic the voting behavior of ensemble learners. We investigate this question by comparing a boosted ensemble of neural networks to a single more complex neural network and providing an intuitive visual example of how the weaker neural networks can collectively learn a stronger hypothesis. Motivated by the positive results of boosting we present a method for partially parallelizing the serial process.

Synthetic Boosting

Our initial focus was to compare and contrast boosting and single more complex networks in a way that could be easily visualized and controlled. To do this we created a synthetic dataset consisting of two intertwined spirals of points representing different categorizations. A boosted neural network ensemble consisted of three detectors each with three layers of 100 nodes and a combined network consisted of three layers of 300 nodes to serve as a single more robust classifier. Additionally, we ran the single detector using three different amounts of data in comparison to the batch size of the boosted detectors: where the single detector shared the same batch size, where it was given three times the batch size, and where it was given 17 times the batch size. These conditions were chosen to give insight as to how the performance of the single network varies as a function of its dataset size in proportion to the ensemble's.

Synthetic Results

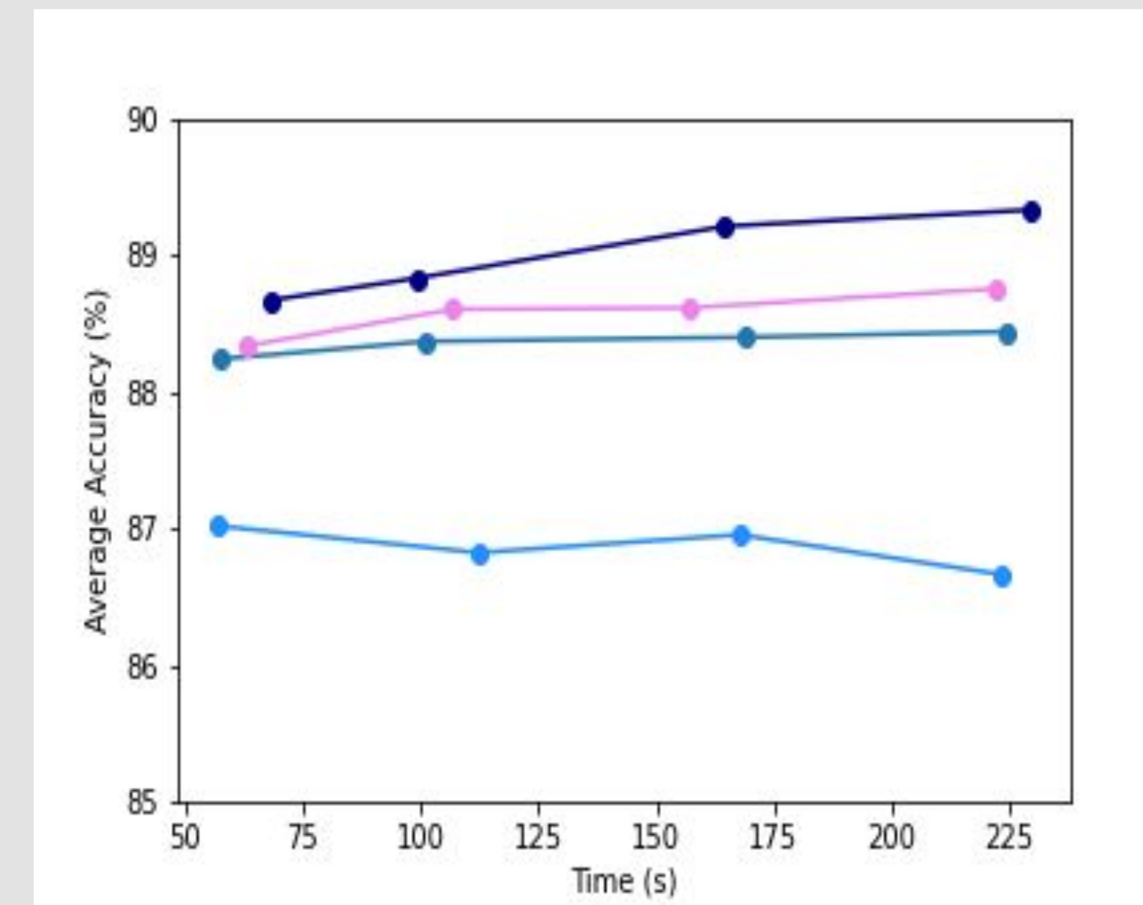


Real World Boosting Motivation

When running the ensemble it was found that the overhead of starting and stopping training along with having to create new datasets added significantly to the run time of the ensemble. To mitigate this effect and provide a more realistic idea of real-world performance we partially parallelized the process of boosting as follows. Instead of waiting for the first and second networks to finish training before creating the necessary datasets, pause training after the detector has gained a reasonable accuracy then use its predictions to form the next dataset so that the detector can finish training while the following detector can get a head start training.

We observed somewhat sporadic behavior in final accuracy in which detectors would sometimes reach an accuracy of over 99% whereas other times they would fall short of 70% which appears to be a function of the highly sensitive initialization state of this classification problem. However, the average accuracy of the ensemble far outperformed its sub networks and showed the potential improvements that can be gained from boosting weak neural networks.

The next dataset we selected was Fashion-MNIST. We found that on this dataset the ensemble outperforms the analogous single network with the same amount of data used for training. It is also reasonable that the network with an significantly larger batch size slightly outperformed the ensemble and that the single network which had less net training data performed worse than the ensemble.



Conclusion

Although boosting is most often discussed in the context of weak learners, it can also be an effective technique when used with advanced learners such as neural networks. Furthermore, a parallelized implementation seems to be able to maintain the accuracy improvements of the conventional serial approach while making its real world run time more competitive with alternative approaches.Much of the behavior of boosting appears to reside in the creation of datasets which emphasize patterns missed by previous detectors. As such the success of boosting can be dependent on how well this dynamic is understood and applied in a given scenario.

Acknowledgement

This work is funded by NSF CCF-1757017 under the Research Experiences for Undergraduates Program