NEURON-SPECIFIC DROPOUT: A DETERMINISTIC REGULARIZATION TECHNIQUE FOR NEURAL NETWORKS

Deep neural networks contain multiple non-linear hidden layers, making them very expressive models that can learn very complicated relationships between their inputs and outputs. With limited training data, however, many of these complicated relationships will be the result of sampling noise. Relationships will exist in the training set, but not in real test data, even if it is drawn from the same distribution. This is known as overfitting. The goal of this project to to fix this problem.

Method

In short, NSDropout works by comparing the outputs of the previous layer for a given batch, and the outputs of the previous layer when the network is passed with a subset of validation data. By averaging the outputs of each neuron for the training and validation data, NSDropout is able to determine which parts of the layer have diverged from the expected behavior of unseen data.

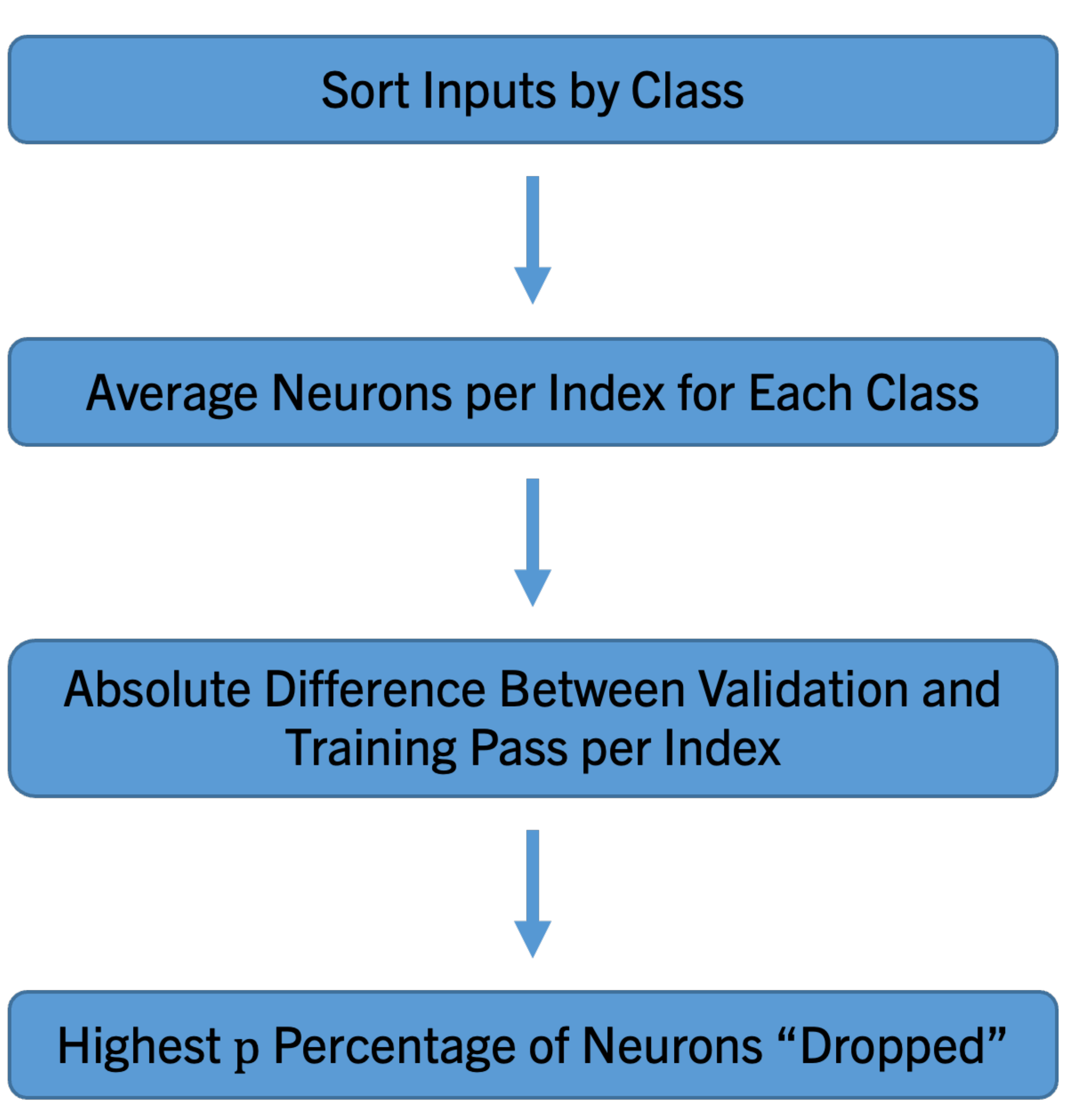

Figure 1. Flow chart of steps for NSDropout.

Related Work

There are many methods that are currently being used to prevent overfitting. Some of the most popular methods include L1 (Lasso Regression), L2 (Ridge Regression), Early Stopping, and Dropout. L1 penalizes absolute value of weights while L2 penalizes sum of squares. NSDropout takes inspiration from dropout in its methods. Dropout randomly drops neurons creating infinitely many networks whereas NSDropout targets specific neurons.

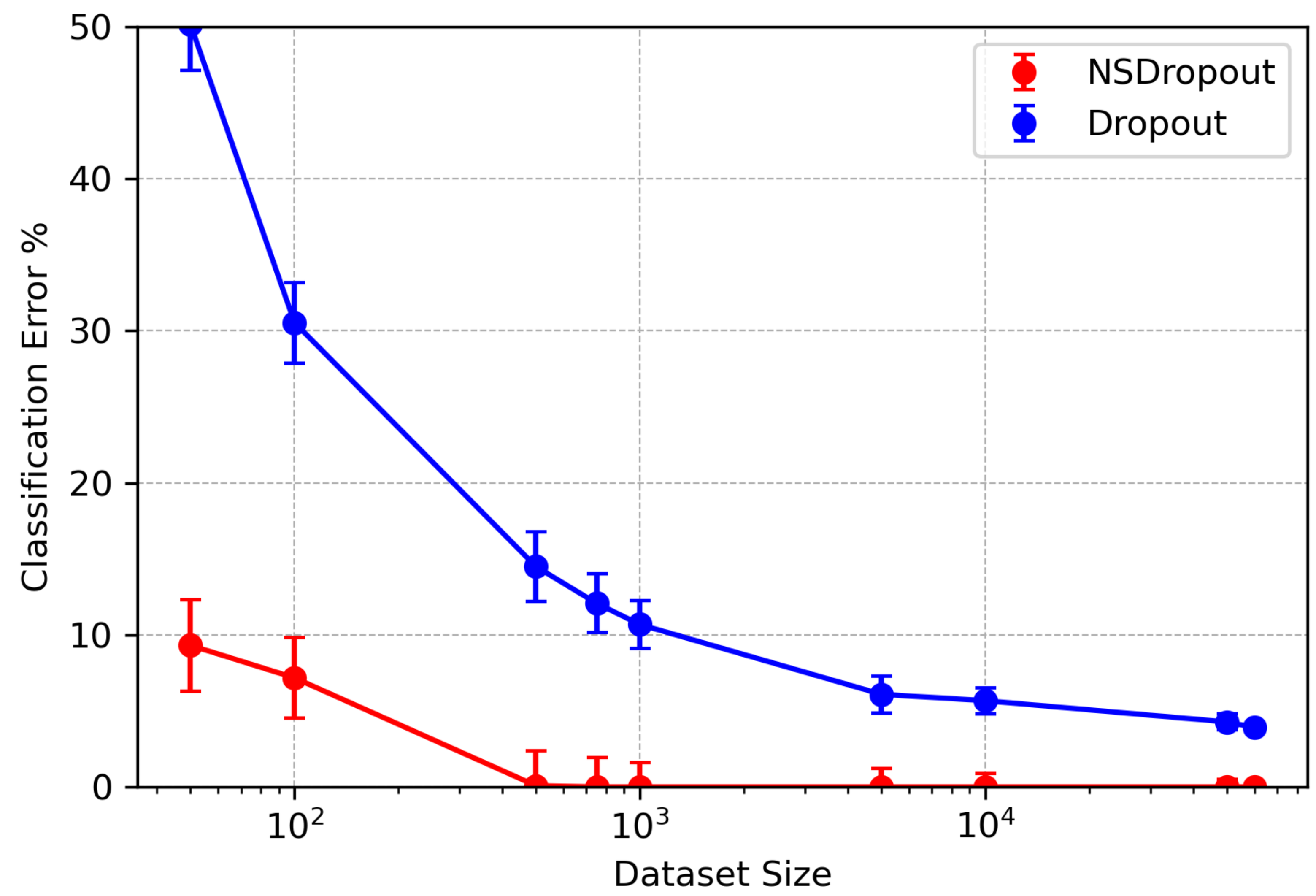

Figure 2. Effect of data set size on error percentage.

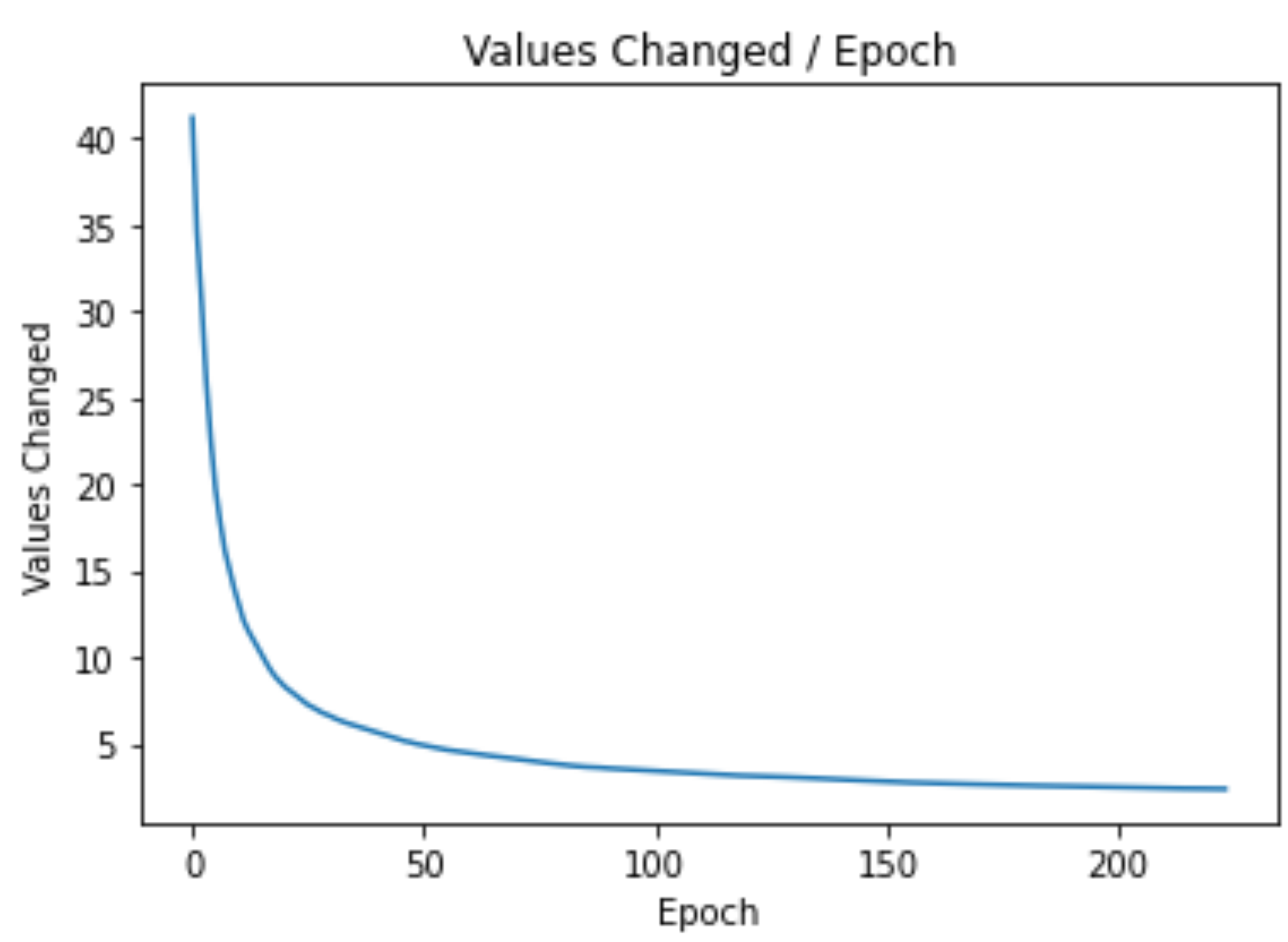

Figure 3: Mask Values Changed from Epoch to Epoch (CIFAR-10 Training).

Discussion

Neuron-specific dropout showed to provide significant improvements compared to traditional regularization techniques. Neuron-specific dropout can be seen as a way to actively prune a model. Neuron-specific dropout, in this sense, is similar to training a model with an actively changing "winning" lottery ticket[1]. It was also found that NSDropout can develop a reliance on the mask given and penalizing this behavior is needed to further development.

Conclusion

Using NSDropout proved to improve the performance of neural networks in image classification domains. NSDropout was able to achieve best-in-class results in MNIST Handwritten Digits, Fashion-MNIST, and CIFAR-10. In addition, to improve the results of image classification networks, NSDropout also reduces the need for large data sets.

References

[1]Frankle, J. and Carbin, M., 2018. The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

All images and figures created by the researcher