Influence of blackout regularization on diverse datasets

Mateusz Garbacz, Michael Schoustra, Ramy Al Sharif, Valerie Pourquie, Thomas Puppels

TU Delft
Delft University of Technology

1 Introduction

The primary task in Machine Learning (ML) is creating a model that generalizes well. Overfitting occurs when the learning model has been attuned too well to the training data at the cost of its generalization performance. There are a few available methods which reduce overfitting:

- Use of more data
- Reduce amount of features
- Reduce complexity/flexibility of the model

The latter two can be considered Regularization methods. L1 regularization operates by deactivating certain weights. However, there is no direct control over the amount of deactivated. We introduce Blackout Regularization (BR), where one can define the percentage of deactivated connections. It is based on L1 and our research will be focused on the influence of blackout regularization on deep learning models.

2 Methods & Workflow

The method is based on the L1 regularization term[2]. In our experiments the amount of regularization applied is controlled with additional parameters.

$$\Omega = \frac{\sum f(w) - \rho}{\rho}\lambda||w||_1$$

ρ : target #active weights
λ : L1 regularization strength
w : weights
$f(w)$: activation function to calculate active weights

In order to achieve the goal, we firstly introduce a function that counts the number of active weights. The calculation is based it on the sigmoid activation function, which we modify to output a number close to 0 if the weight is around 0 and close to 1 otherwise (Fig. 2).

In order to control the amount of the L1regularizationadded to the loss function we measure the relative differenceof the actual number of active weights from the desired onep:

The method has two parameters to optimize: the target number of active weights ρ (input by the user as thepercentage of all weights available in the network) and λ.

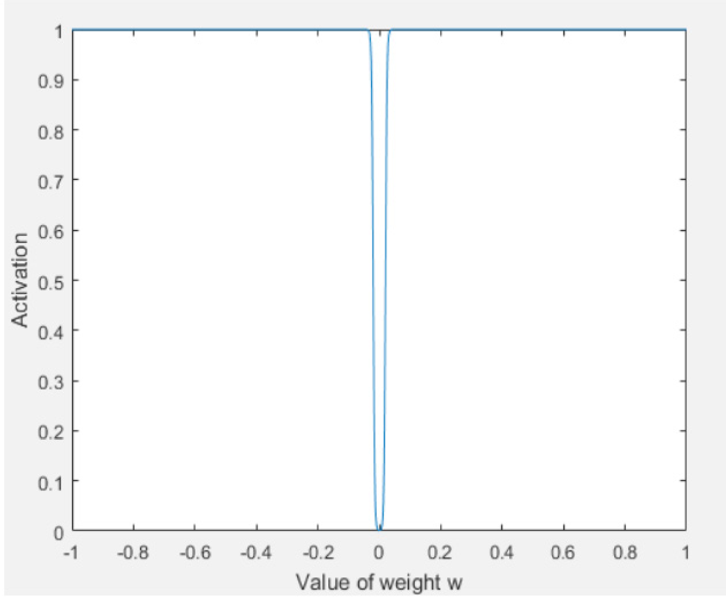

Figure 2: The activation function used to calculate the whether the weight is active.

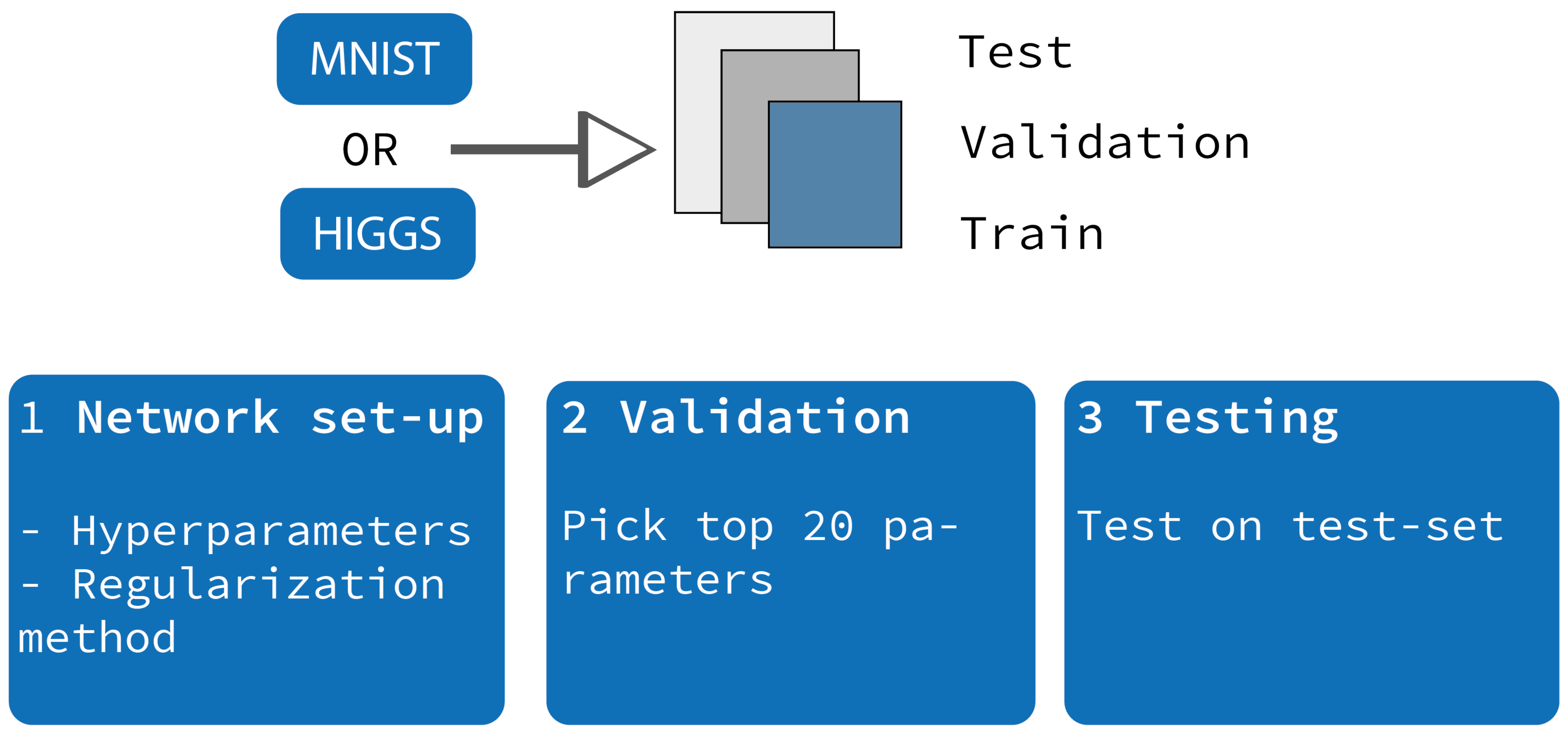

Figure 3: Work flow. For each dataset, and each regularization setting per dataset,a network with parameters defined by the random search istrained on training data and tested on validation data. Fi-nally, the top 20 performing parameter settings per regular-ization are tested with test data.

3 Experiments

The main questions is "**How does Blackout Regularization influence deeplearning mode?**". To answer this question. the research is divided into four subquestions.

1. Does BR prevent overfitting, and if so, how well does it so?

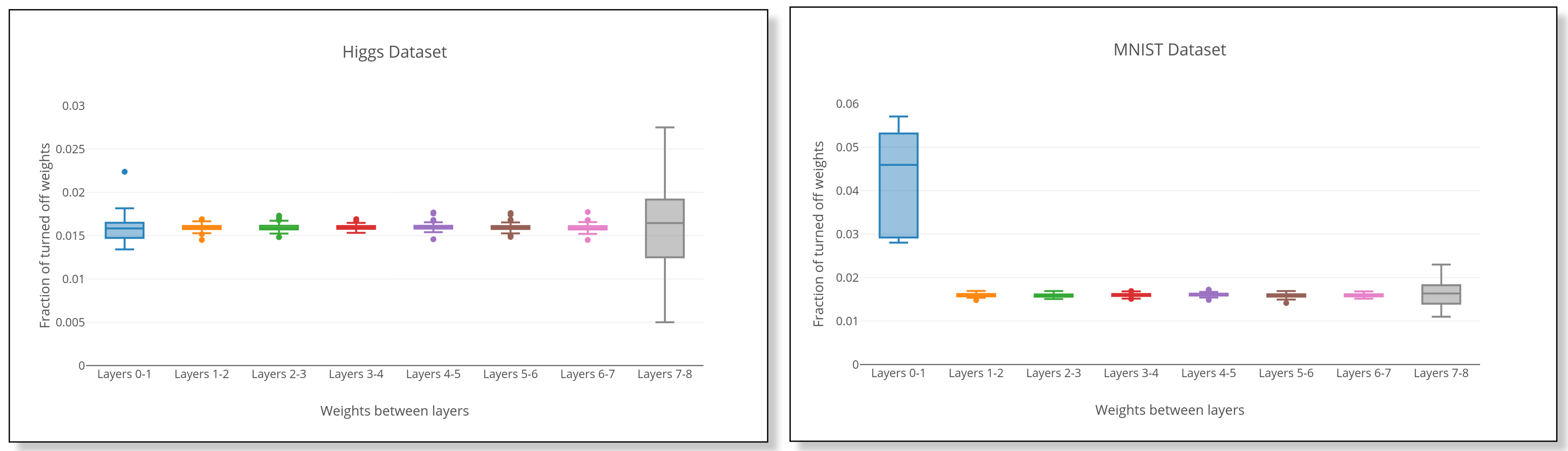

Figure 4. The test and validation accuracy for the MNIST and HIGGS dataset plotted for different regularization types and without regularization. The blackout regularization has a small effect on the MNIST dataset but seems to have a negative influence for the HIGGS dataset.

2. Which layer's connection weights are most often driven to zero?

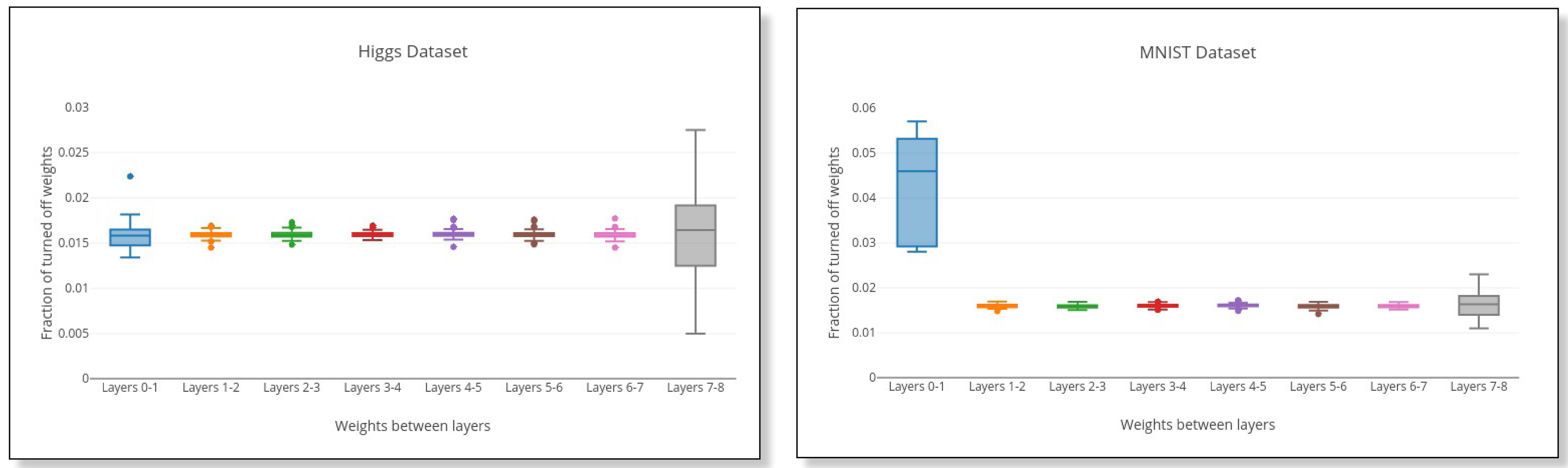

Figure 5. Which layers is turned vs the percentages of deactivations. For the MNIST the first layer is turned of most often, for the HIGGS dataset the amount of times each layer is deactivated is similar for all layers.

3. How does dataset-size influence the results of a NN using BR?

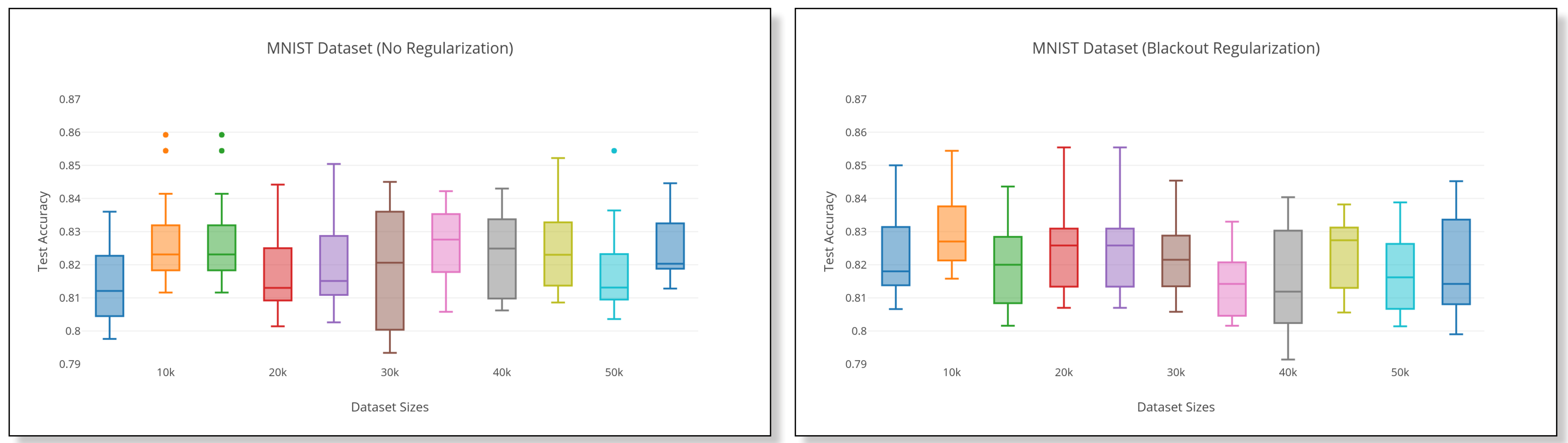

Figure 6. The dataset size vs the test accuracy for the MNIST dataset. From the figures it can be seen that for larger datasets, the blackout regularization has a negative effect while for small datasets it has a positive influence.

4. How does the blackout parameter influence the performance?

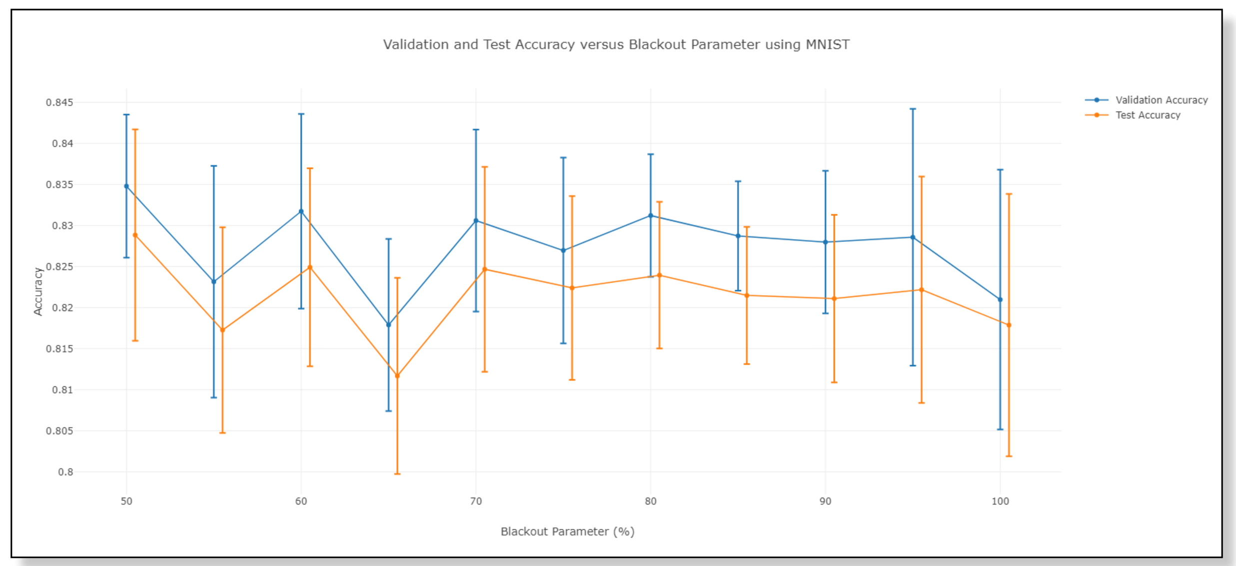

Figure 7. The dataset size vs the test accuracy for the MNIST dataset. From the figures it can be seen that for larger datasets, the blackout regularization has a negative effect while for small datasets it has a positive influence.

4 Discussion & Conclusion

References

[1] S. Scardapane, D. Comminiello, A. Hussain, and A. Uncini. Group sparse regularization for deep neural networks. Neuro-computing, 241:81 – 89, 2017.
[2]Y. B. Ian Goodfellow and A. Courville. Deep learning. Book in preparation for MIT Press, 2016.