Supplementary Material

for the report

Density of States of Neural Network

Constantin Pinterits

1 Additional Details on Dataset variation

As mentioned in Section 5 of the report, it has been observed, that the Density of States can vary drastically for low and high energy ranges. This is especially true for the single layer perceptron architecture. Figure 1 shows three unique simulations performing a binary classification task on the first two classes of the MNIST Dataset by a single layer perceptron. The Hyperparameters are set to $P = 140$ and $N = 100$ with a balanced Dataset $P_1 = P_2$. It is easy to see, that the height of the peak changes with each simulation and that the bulges in the low and high energy regions are more or less pronounced. For every simulation a different Dataset was used which not only had an influence on the output of the Wang-Landau Algorithm but also on the computation time of the code. The light green graph took 9 minutes to converge, while the dark green graph ran for almost 130 minutes. In light of this fact, it seems, that the Dataset represented in the latter simulation consists of images that are difficult for the network to label correctly. As the graph shows, it is not very likely for the system to find a configuration that correctly labels all 140 images. This could be one of the reasons why the discrepancy between Figure 4a and Figure 4b arise, but other explanations like a different initialization or a varied sampling method might also play a roll.

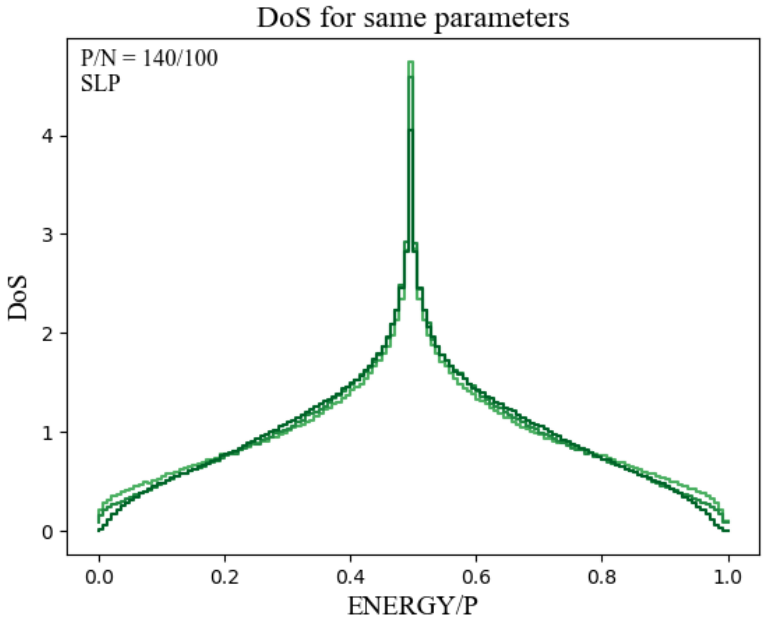

Figure 1: DoS for the same Hyperparameters