Exercise 2: Backpropagation and Feedback Alignment

Julian Büchel, 16-943-938, jubueche@ethz.ch

September 2020

1 2.2.1 Linear case

Since the implementations of backpropagation and feedback-alignment are equal if $B = W^T$ one would expect that if B converges towards W^T (meaning the angle goes to 0), performance also approaches the performance of backpropagation (because then backpropagation is essentially implemented). The figure below shows this property.

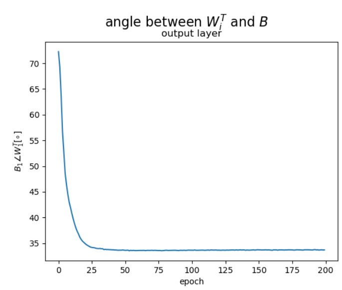

Figure 1: Linear: Angle between B and W_i^T for the output layer.

1.1 [Optional] Does it need to send useful feedback signals back to the layer?

Since the hidden layer has only 20 units compared to the 30 input units it would not be sufficient to learn the output matrix) $\in \mathcal{R}^{10\times 20}$ given random feed-forward weights into the hidden layer. If the hidden layer would have the same number of inputs and the input matrix would have full rank, then it would suffice to only learn the output weights.

2 2.2.2 Nonlinear case

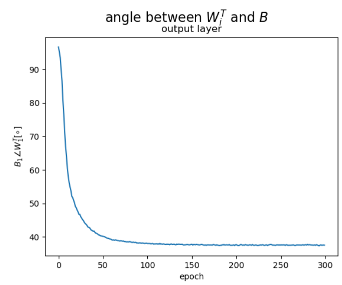

Figure 2: Non-linear: Angle between B and W_i^T

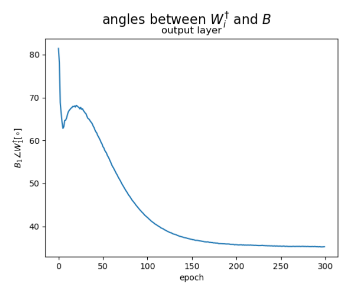

Figure 3: Non-linear: Angle between B and $W_i^\dagger$

2.1 Does it also work for non-linearities?

Since the angle between the transpose and the random feedback weights also starts to align I think that FA also works in the non-linear case, but (as we heard in the lecture) not as well as back-propagation.

2.2 Compare with backpropagation

For the same number of parameters, backpropagation performs better with a test MSE of 0.007. FA performs worst with a test MSE of 0.012.

One obvious reason is that the feedback weights B do not align with the true inverse of the weights W, leading to slight incorrectness in the computation of the gradient and therefore worst performance.

2.3 2.2.3. Polynomial fitting

FA: Test MSE: 1.39 BP: Test MSE: 2.48

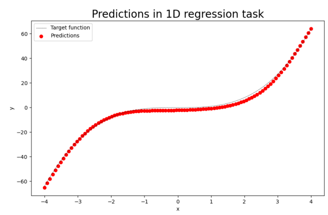

Figure 4: Polynomial regression using FA

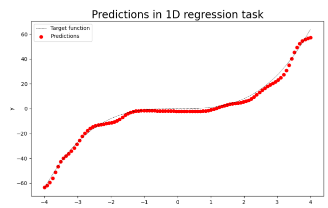

Figure 5: Polynomial regression using BP

Although the training MSE of the model trained with BP was lower, the test MSE of the model trained with FA was lower. This is a clear sign that the BP model overfitted the data, which can also be seen in the plot. The reason for this is simply that the model has a too high complexity. The model trained with FA does not suffer from this. The reason is not that the model trained with FA is "robust" to over-fitting, no, it just did not minimize the error to such a high degree that overfitting started to happen.