Model Pruning: Weekly Report 4

Patricia Gschoßmann

1. Weekly Progress

In this week selected experiments from last week were continued to improve the result. As before, the goal was to prune 65% of the parameters at once. The following changes were applied compared to last week:

- Initialize weights of convolutional layers in each parallel branch appropriately using He initialization[1].

- Remove early stopping during α-schedule

- Train for a maximum of 350 epochs for each α

- Reduce the initial learning rate from 0.1 to 0.001

- Replace the previous learning rate scheduler; reduce the learning rate, whenever the validation loss stops improving[2]

- Regarding regression: The loss now corresponds to the average MSE between each output value and its target (instead of the sum)

2. Results

2.1. Classification

In last week's experiments it was observed that the training for different α-values stopped before the model reached its final performance, because of the early stopping callback. Figure 1 shows intermediate results for the new set up without this callback for different α-values between 1 and 0.4.[3] The same α-schedule as last week was used. So far, the model maintains its original training accuracy. Moreover, the model reached a higher validation accuracy for lower α-values as expected due to a longer training period. What is also interesting is, that model with lower α-values sometimes outperform higher α-values.

An improvement compared to last week is, that so far the model reached a validation accuracy higher than 90% for each α-value. The experiment is still running, but based on

this observation it is more likely, that the models maintains its performance until $\alpha = 0$.
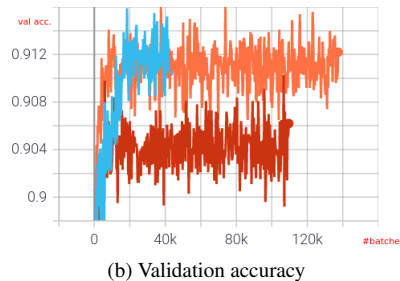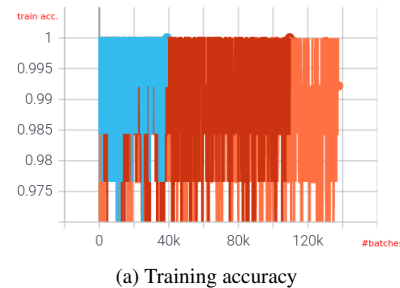

(a) Training accuracy

(b) Validation accuracy

Figure 1: Development of train- and validation accuracy during training for $\alpha = 0.95$ (orange), $\alpha = 0.6$ (red), $\alpha = 0.45$ (light blue). An exponential schedule with a decay rate of 0.05 was used.
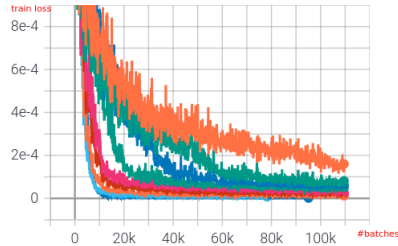
2.2. Regression

Figure 2 shows the training and validation loss for different α values between 1 and 0.6 for the new set up using regression. The same α-schedule was used as last week. The development of the training loss clearly improved compared to last week: It decreases continously for each α. It can also be seen that the loss would continue to decrease if the training units were extended to more than 350 epochs. However, investigating the graph of the validation loss shows reveals extremly overfitting, which is why this experiment was aborted early. Further investigation is needed to discover the reason for this behavior. Independently of that observation, an alternative approach would be to optimize over each logit error separately instead of optimizing based on their average error.
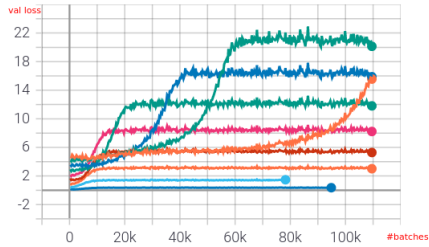
---

[1] https://openaccess.thecvf.com/content_iccv_2015/papers/He_Delving_Deep_into_ICCV_2015_paper.pdf

[2] To clarify: The learning rate is reset to 0.001, whenever α drops, i.e. the learing rate schedule restarts.

[3] The experiment is still running.

(a) Training loss

(b) Validation loss

Figure 2: Train- and validation loss during training as regression task for α-values between 1 and 0.6. An exponential schedule with a decay rate of 0.05 was used.
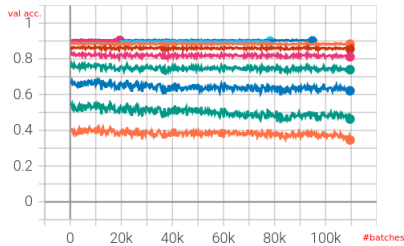

Figure 3: Development of validation accuracy during training as regression task for α-values. An exponential schedule with a decay rate of 0.05 was used.

3. Plan

In the following week I continue with my current approach for the classification task, based on the (still open) results for the currently running experiment. Moreover I want to investigate following approaches in order to improve my results:

- Debug regression loss

- Allow weight updates for the pretrained model.