

# 1 Exercise 1 - Eager Mode

## 1.1 Saturation Plots

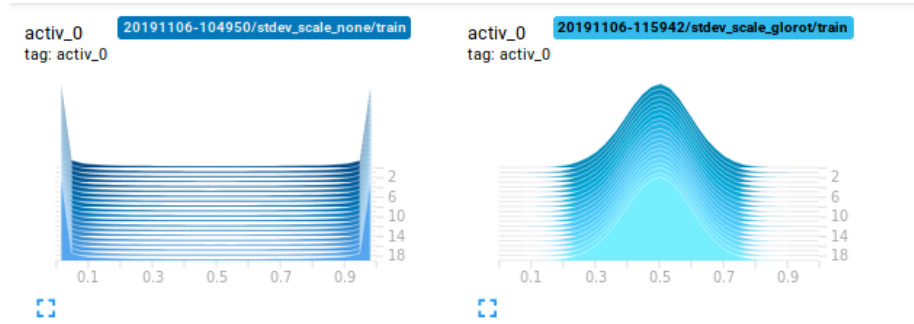


Figure 1: Activations after weight initialization without scaling (left), and Glorot initialization in layer 0(right)

Looking at Figure 1 we can clearly see that the initialization without standard deviation scaling is in the saturated region. However the Glorot initialization seems to work well.

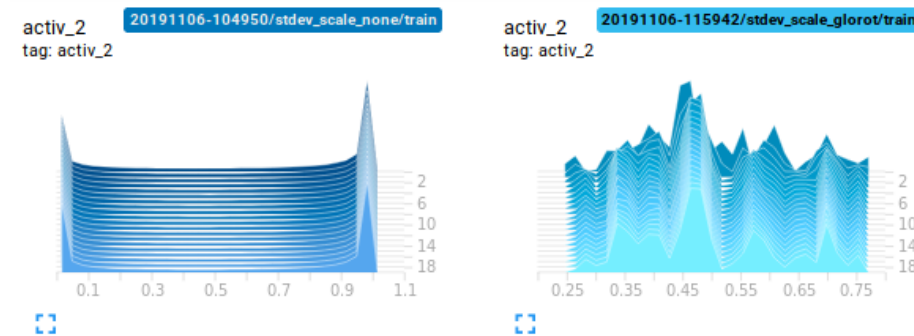


Figure 2: Activations after weight initialization without scaling (left), and Glorot initialization in layer 2(right)

After skipping layer one as it looks more or less similar to layer 0, especially for the left side, we added a plot of layer 2 (Figure 2). We clearly see that the activations are becoming well distributed over the whole sigmoid "range".

In layer 3 the activations are saturated for the model without glorot initialization, but have high peaks for the model with glorot initialization (Figure 3). We might conclude that this model seems to be "sure" about its prediction whereas the other is less sure.

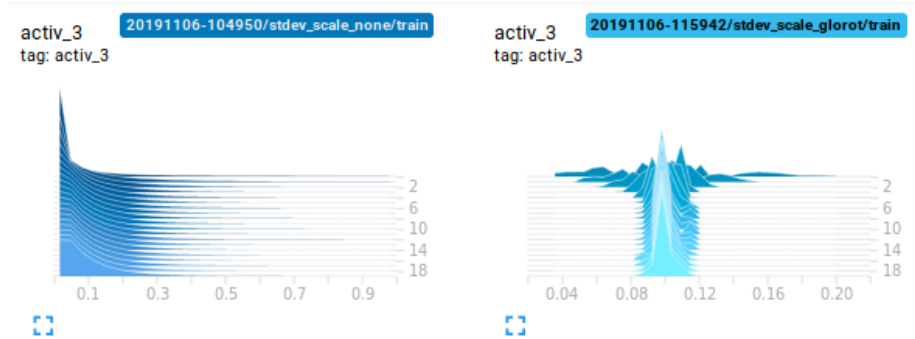


Figure 3: Activations after weight initialization without scaling (left), and Glorot initialization in layer 3(right)

## 2 Exercise 2 - Keras model

The models with following different settings (the shared parameters are not noted here):

- Reference Model without any dropout Layers
- Dropout Layer after each hidden layer except the first
- Dropout layer and adaptive learning rate depending on the epoch (See also Figure 6).

The models are compared in Figure 4. We can clearly see that dropout layers and adaptive learning rates lead to better results.

In Figure 5 the loss values w.r.t. models are plotted. Note that the model with dropout layers but without adjusted learning rate was only at the beginning of convergence whereas the reference model and the one with adaptive learning rate converges nicely. Especially the one with adaptive learning rate converges faster than the other models. It might be that the model with dropout layers but without adaptive learning rate fluctuates around a cost function minima.

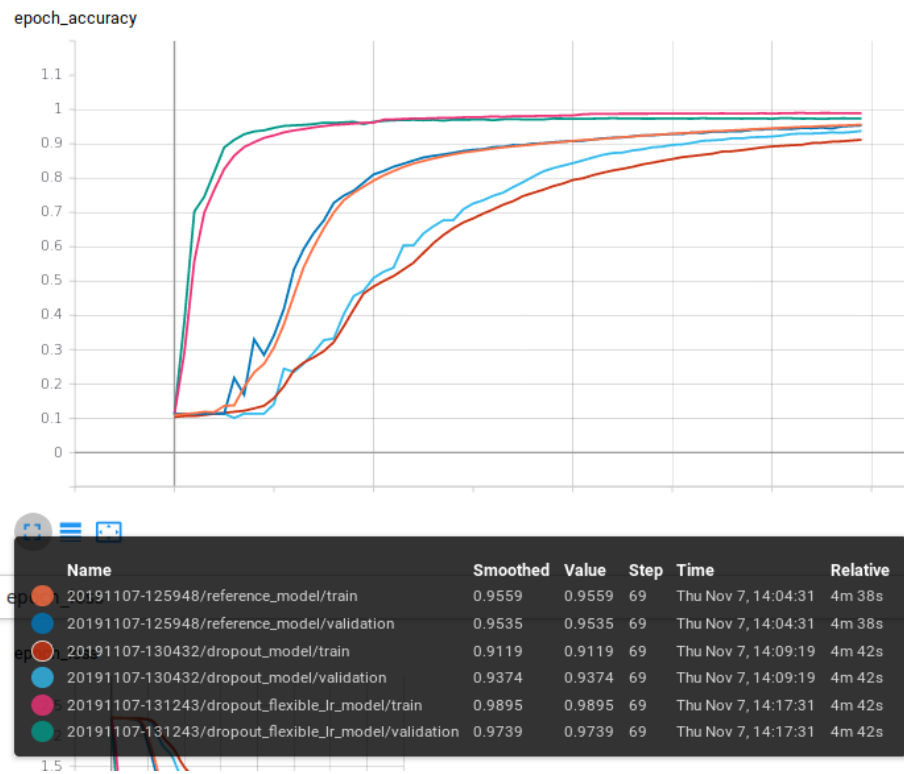


Figure 4: Model accuracy comparison. Note that the model with dropout layer and adjusted learning rates is fairly the most competitive.

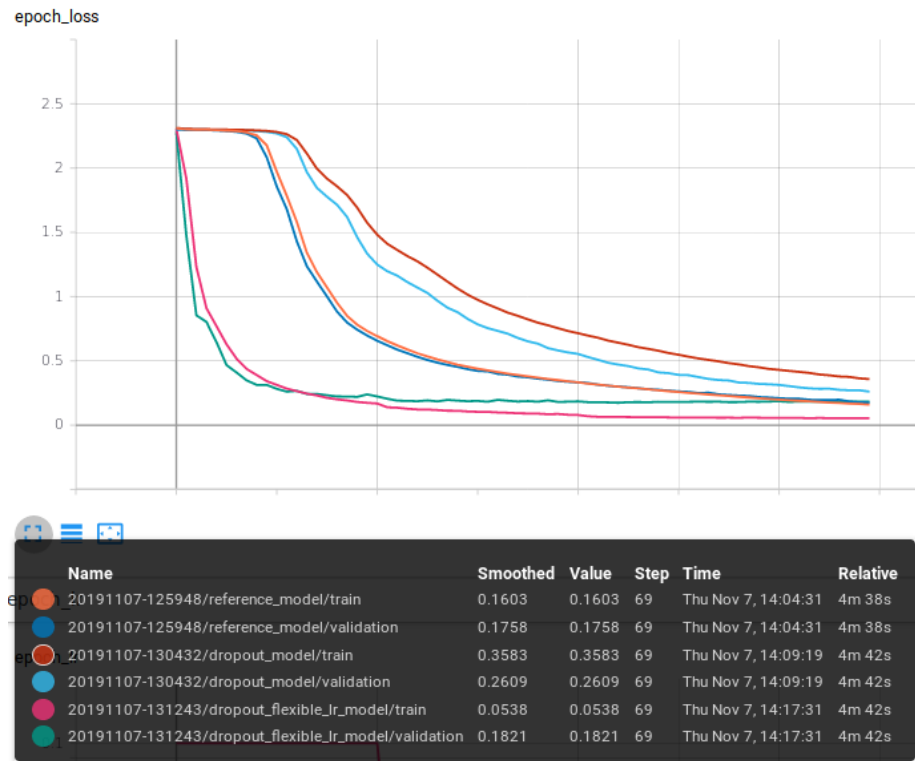


Figure 5: Loss values in comparison. One can observe that the validation-set for the model with dropout layers and flexible learning rates is on the loss value of the reference model.



Figure 6: Learning rate for the third model. The learning rate was 0.1, then was dropped to 0.05, afterwards to 0.01. It was implemented using a keras *LearningRateScheduler*.