Exercise Sheet Deep Learning
Part1: Foundations of Deep Networks
Summer 22

This sheet includes a theoretical part and a practical assignment on the first part of the lecture Deep Learning (1_Foundations). Both parts give 20 points maximum each. Please hand in solutions as a pdf in groups of at most three persons via LernraumPlus.

name1:

name2:

name3:

PARTI – THEORY: For the following, you might answer only YES/NO (or abstain), or you can add short arguments (at most two lines per question). If you are not sure, it is better to abstain.

1. The following arguments constitute valid reasons why one could prefer training of deep networks rather than shallow ones:

yes **no** Deep networks are universal approximators, but shallow ones are not.

yes **no** Minimizing the error of only one single neuron is already NP-hard, so from a theoretical perspective it is not much more costly to train deep networks. In both cases, in particular, numeric optimization method based on gradients work in a similar way.

yes **no** Overparameterization and long training times can lead to the so-called double descent effect, i.e. a region where deep networks do not overfit but the generalization error decreases.

yes **no** Deep networks can more efficiently realize highly nonlinear problems, and there exist proofs of this, such as a quantification of the topological complexity of their decision surface.

2. In deep network training, the following twists are often used in training for better numerical behavior or generalization ability:

yes **no** change of activation function to the one abbreviated as SeLU;

yes **no** change of model architecture to convolutional ones for image processing;

yes **no** Monte Carlo dropout;

yes **no** invertibility of models;

3. The following network architectures have the specific property that ...

yes **no** residual models are invertiblae

yes **no** U-net architectures can directly deal with images of different size

yes **no** Siamese twin architectures enable few shot learning, i.e. learning new classes based on one example only

yes **no** The deep adaptation network also tries to keep the distribution of hidden layer activations the same as for the original training data

4. The following networks from the model zoo approach the task:

yes **no** YOLO for object detection

yes **no** OpenPose for Protein Folding

yes **no** Inception for visual question answering

yes **no** EfficientNet for image classification

5. The following is true:

yes **no** U-net architectures cannot be used together with convolutions

yes **no** Fully convolutional architectures can deal with inputs of different sizes

yes **no** A valid pooling operator is the concatenation of all values as a vector

yes **no** ADAM refers to the variation of SGD with normalized step size per batch.

PARTII – PRACTICE: You can use code and models which are publicly available, please clearly reference such sources. It might be a good idea to start with the examples given in the practical part of the lecture (available at `https://jgoepfert.pages.ub.uni-bielefeld.de/talk-deep-learning`). Please give a link to your code, and please describe the experiments and results of your approach in a pdf which is well structured (e.g. modeling/training parameters/training/results/interpretation, use itemize, keywords are fine) and enables reproducability as well as easy access to your main results . Please use at most one page for both practical parts together including graphs and images.

1. Take the Fashion MNIST data set and a convolutional network architecture. Display the difference of the two optimization algorithms ADAM and SGD w.r.t. training and test error.

2. Take the same setup and the ADAM optimizer, display the result of the choice of different batch sizes for mini-batch training. Take at least three choices and shortly discuss the differences.