Barbara Hammer, Philip Kenneweg Universität Bielefeld

Deep Learning (SoSe 2024)
3. Sheet

Start: Thursday, 23.05.2024.
End: The worksheets should be solved using Python, in groups of 2-3 people and will be presented in the Tutorials.
Discussion: Thursday, 13.06.2024 in the Tutorials.

Information

The worksheets and necessary toolboxes will be made available in the Lernraum "392221 Deep Learning (V) (SoSe 2024)". Worksheets will usually be released every two weeks on Thursday, and discussed during the exercises on Thursday two weeks later. In order to successfully finish the course, 50% of the available points have to be obtained and each participant has to present his/her results at least once. The Wednesday and Thursday in between the release and discussion of the sheet will be used to discuss the implementation of the various algorithms presented in the lecture, as well as go deeper into the relevant material.

Exercise 1:

(*10 Points*)

You can use code and models which are publicly available. Please provide: short description what you did, how it is done, what is the result. Please be prepared to present the solution in the exercises (best in form of a Jupyter notebook .ipynb).

(a) *(5 Pts.)* Use a classification deep network for the MNIST [1] data set. Perform at least three different types of targeted attacks on 5 different numbers (2 pts), including one attack which puts particular effort on the fact that the attacked pattern is indistinguishable from the original one. (1 pts) Evaluate the performance of the attacks visually (1 pts) (which attack does not change the visual impression) and quantitatively (1 pts)(distance of attack to original sample, success rate of the approach).

(b) *(5 Pts.)* Use the FashionMNIST [2] data set and a deep model. Create a universal attack, which works for multiple different inputs. Describe how you approach this, and evaluate the performance (success rate). (3 pts) Evaluate whether the universal attack also transfers to other deep learning architectures. (1 pts) Can you also create a universal attack which is hard to detect visually? (1 pts)