

Deep Learning for NLP 2021

Homework 1

Group 50 - David Kösters und Jonas Rudolph



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Summer semester 2022

2. Sigmoid Activation Function

3.1 Circular Dataset

Although single-layer perceptrons are only capable of learning linearly separable patterns, with the right input features the perceptron is able to learn a good discriminator for the circular dataset in the TensorFlow Playground. But in the given Figure with data point distributed in a 2-dimensional space it shouldn't be possible to discriminate the two classes.

3.2 MLP

Increasing the number of neurons from hidden layer to hidden layer towards the output seems to result in the worst results and slowest convergence. Although an equal number of neurons in each hidden layer leads to better results, placing more neurons towards the input a decreasing the number of neurons in each consecutive hidden layer results in the fastest convergence and produces the best results. This might be due to the fact that some input information will get lost when placing too little neurons towards the input of the network as well as neurons not having a chance to focus on and „learn“ specific key features which are needed to predict the labels accurately.

4. Softmax

Because the softmax activation function utilizes an exponential normalization (e.g. to represent the relative probability of an output), the function is not applied locally at one neuron but rather incorporates information from all neurons in a given hidden layer.