Deep Learning – MLP

Summer 2023
Prof. Dr. Jochen Schmidt

Technische Hochschule Rosenheim

1) MLP-Parameters

You design a fully connected neural network with 4 hidden layers, each with 10 units. The input is 15 dimensional, the output is a scalar. All activations are sigmoids.
What is the total number of trainable parameters in your network? 15, 10, 10, 10, 10 1

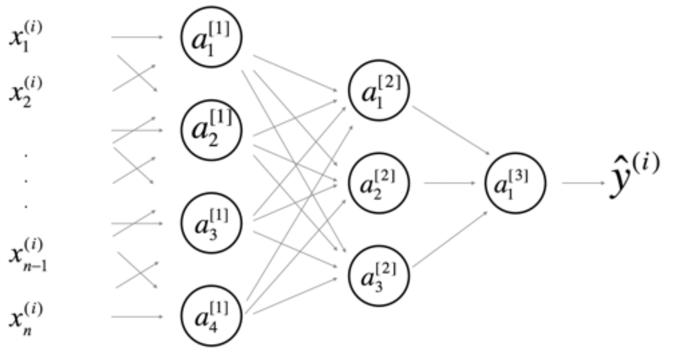
150 + 100 + 100 + 100 + 10
= 460 + 41 bias = 501 total

2) Universal Approximation

Recall that a neural network with a single hidden layer is sufficient to approximate any continuous function (with some assumptions on the activation). Why would you use neural networks with multiple layers? A MLP can reach the same model quality with less parameters and/or converge faster.

3) MLP-Training

You design the following 2-layer fully connected neural network. All activations are sigmoids and your optimizer is gradient descent.



a) You initialize all the weights and biases to zero and forward propagate an input $\boldsymbol{x}$ in the network. What is the output $\hat{y}$? ^y = sig(0) = 0.5

(a) –1 (b) 0 (c) 0.5 (d) 1

b) Consider the model with all parameters initialized with zeros. $\boldsymbol{W}_1$ denotes the weight matrix of the first layer. You forward propagate a batch of examples, and then backpropagate the gradients and update the parameters. Which of the following statements is true?

(a) Entries of $\boldsymbol{W}_1$ may be positive or negative
(b) Entries of $\boldsymbol{W}_1$ are all positive
(c) Entries of $\boldsymbol{W}_1$ are all negative
(d) Entries of $\boldsymbol{W}_1$ are all zeros

change: sig(x) * sig'(x) * x
= 0.5*(0.5*0.5) * x
pos or neg, depending on x

c) Consider the model with all parameters initialized randomly with large positive numbers. Is this a good idea? no, this way sigmoid will output a lot of high values and take a very long time to train, if it ever will converge.

Acknowledgements: Exercises based on course CS230 “Deep Learning”, Andrew Ng et al., Stanford University.