EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATH. - NATURWISS. FAKULTÄT
FACHBEREICH INFORMATIK
KOGNITIVE SYSTEME · PROF. A. ZELL

Deep Neural Networks

Assignment 9

Assignment due by: 10.07.2018, Discussions on: 17.07.2018

Question 1 Backpropagation by Hand (6 points)

For the same neural network on slides 73/74, perform a pass of backpropagation, calculating the gradients w.r.t. the weights and the biases numerically with the target value $\hat{\boldsymbol{y}} = [0, 1]$. Please note that in this example the matrices are multiplied like: $\boldsymbol{a}^{(1)} = \boldsymbol{W}^{(1)}\boldsymbol{x} + \boldsymbol{b}^{(1)}$, so the matrices are shaped accordingly (instead of $\boldsymbol{x}\boldsymbol{W}$ like in previous exercises).

Question 2 Symbol to Symbol derivatives (3 + 4 points)

For this question (and question 3), we will use the following network:

$$\mathcal{L} = \frac{1}{2} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2$$

$$\boldsymbol{y} = \boldsymbol{u}_3 + \boldsymbol{z}_2$$

$$\boldsymbol{u}_3 = \boldsymbol{W}_3 \boldsymbol{z}_1$$

$$\boldsymbol{z}_2 = f(\boldsymbol{u}_2)$$

$$\boldsymbol{u}_2 = \boldsymbol{W}_2 \boldsymbol{x}$$

$$\boldsymbol{z}_1 = f(\boldsymbol{u}_1)$$

$$\boldsymbol{u}_1 = \boldsymbol{W}_1 \boldsymbol{x}$$

where $\boldsymbol{x}$ is the input, $\hat{\boldsymbol{y}}$ is the target output and f is the activation function.

(a) Convert this neural network into a $\mathbf{directed}$ graph (like those in pp. 65-67 & 79). The input, variables, the loss and intermediate results should be nodes and (directed) edges should connect a node to other nodes that use its result directly.

(b) To this graph add all the nodes required to compute the gradients w.r.t. the weights. In this part, you can leave out any node that has just one input and one output otherwise the graph will get too messy. Please differentiate the nodes that you add in this part from the nodes of part (a).

Question 3 Backpropagation by Hand (7 points)

Perform both the forward and backward pass for the network from question 2 and compute the gradients w.r.t the weights. Please show all the intermediate calculations as well. The values for the inputs and variables are as follows:

$$\boldsymbol{x} = [1, 2, 3, 4]^\top$$

$$\hat{\boldsymbol{y}} = [1, 0]$$

$$\boldsymbol{W}_1 = \begin{bmatrix} 0.4 & 0.5 & 0.25 & 0.1 \\ 0.2 & 0.25 & 0.4 & 0.5 \\ 0.1 & 0.75 & 0.3 & 0 \end{bmatrix}$$

$$\boldsymbol{W}_2 = \begin{bmatrix} 0.1 & 0.25 & 0.5 & 0.4 \\ 0 & 0.3 & 0.75 & 0.1 \end{bmatrix}$$

$$\boldsymbol{W}_3 = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.2 & 0.3 & 0.4 \end{bmatrix}$$

The activation function f for this network is ReLU.

The numbers in this exercise aren't as nice as in question 1, so you can round to two decimal digits in intermediate steps if you want.