

Home Exercise 03

Viet-Hoa Nguyen

Martikeldnummer: 2581721

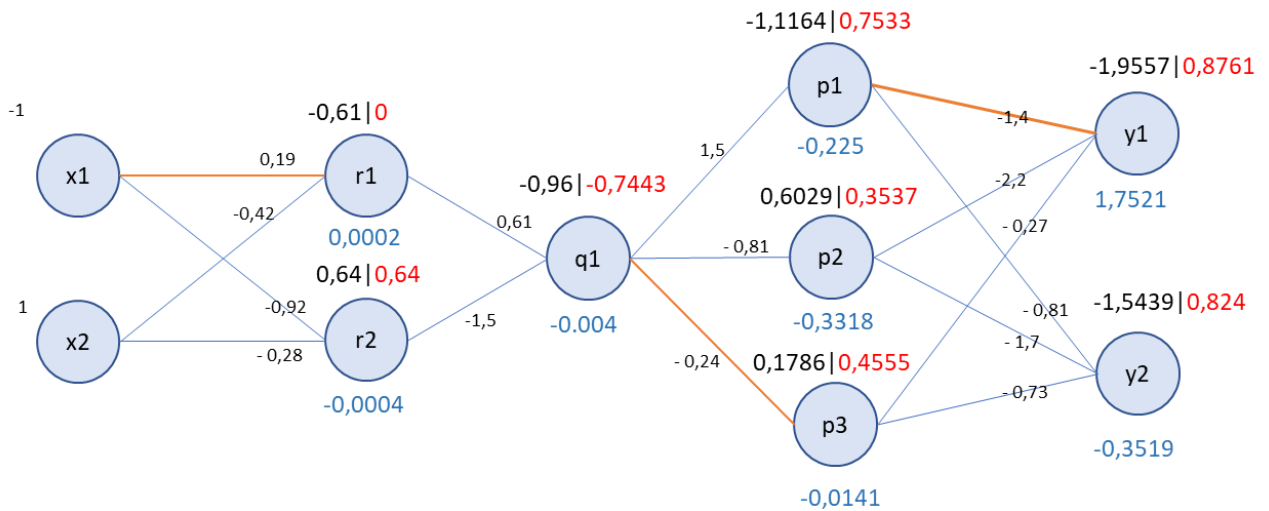
20-00-0947 Deep Learning for Natural Language Processing

Technische Universität Darmstadt

11. May 2020

Problem 1: Backpropagation by hand

Color code: Preactivation, **Activation**, **Error derivative**



1. Forward pass

$$z_1 = X \cdot T = \begin{bmatrix} -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0,19 & -0,92 \\ -0,42 & -0,28 \end{bmatrix} = \begin{bmatrix} -0,61 & 0,64 \end{bmatrix}$$

$$a_1 = \text{relu}(z_1) = \begin{bmatrix} 0 & 0,64 \end{bmatrix}$$

$$z_2 = a_1 \cdot U = \begin{bmatrix} -0,61 & 0,64 \end{bmatrix} \cdot \begin{bmatrix} 0,61 \\ -1,5 \end{bmatrix} = -0,96$$

$$a_2 = \tanh(z_2) = -0,7443$$

$$z_3 = a_2 \cdot V = -0,96 \cdot \begin{bmatrix} 1,5 & -0,81 & -0,24 \end{bmatrix} = \begin{bmatrix} -1,1164 & 0,6029 & 0,1786 \end{bmatrix}$$

$$a_3 = \text{sigmoid}(z_3) = \begin{bmatrix} 0,7533 & 0,3537 & 0,4555 \end{bmatrix}$$

$$z_4 = a_3 \cdot W = [0,7533 \quad 0,3537 \quad 0,4555] \cdot \begin{bmatrix} -1,4 & -0,81 \\ -2,2 & -1,7 \\ -0,27 & -0,73 \end{bmatrix} = [-1,9557 \quad -1,5439]$$

$$a_4 = \text{sigmoid}(a_4) = [0,8761 \quad 0,824]$$

2. Backpropagation

$$\frac{\partial E}{\partial y} = \begin{bmatrix} \frac{\partial E}{\partial y_1} \\ \frac{\partial E}{\partial y_2} \end{bmatrix} = \begin{bmatrix} 1,7521 \\ -0,3519 \end{bmatrix}; \quad \frac{\partial E}{\partial p} = \begin{bmatrix} -0,225 \\ -0,3318 \\ -0,0141 \end{bmatrix}; \quad \frac{\partial E}{\partial q} = [-0,0004]; \quad \frac{\partial E}{\partial r} = \begin{bmatrix} 0,0002 \\ -0,0004 \end{bmatrix}$$

Result:

$$\begin{aligned} \frac{\partial E}{\partial w_{1,1}} &= 1,7521 \cdot \text{sig}(-1,9557) \cdot (1 - \text{sig}(-1,9557)) \cdot 0,7533 = 0,1433 \\ \frac{\partial E}{\partial v_{1,3}} &= -0,0141 \cdot \text{sig}(0,1786) \cdot (1 - \text{sig}(0,1786)) \cdot (-0,7443) = 0,0026 \\ \frac{\partial E}{\partial t_{1,1}} &= 0 \end{aligned}$$

Problem 2: Sentiment Polarity in Movie Reviews

2.2 MLP in Tensorflow

Test result for baseline model with activation function in the order relu, tanh, sigmoid.

Average mse loss over all epoch: 0,47060

Accuracy: 44,58%

2.3 Hyperparameters optimization

A. Known from the previous exercise

Could not find parameters that beat the baseline in 2.2

B. New parameters:

Learning rate: 0.001	Loss: 0,3180
Epoch: 1000	Accuracy: 30,08%
Hidden Layer 1: 30 neurons	Hidden Layer 1: 10 neurons

Learning rate: 0.001	Loss: 17,64
Epoch: 50	Accuracy: 40,05%
Hidden Layer 1: 50 neurons	Hidden Layer 1: 50 neurons
With L2 alpha = 0,01	

Learning rate: 0.001	Loss: 0.47546
Epoch: 50	Accuracy: 44.93%
Hidden Layer 1: 50 neurons	Hidden Layer 1: 50 neurons
With Adagrad	MSE Loss

Learning rate: 0.001	Loss: 17,64
Epoch: 50	Accuracy: 43,08%
Hidden Layer 1: 50 neurons	Hidden Layer 1: 50 neurons
With Adam	MSE Loss

Learning rate: 0.001	Loss: 17,64
Epoch: 50	Accuracy: 40,05%
Hidden Layer 1: 50 neurons	Hidden Layer 1: 50 neurons
With L2 alpha = 0,01	MSE Loss

Learning rate: 0.001	Loss: 17,64
Epoch: 50	Accuracy: 43,38%
Hidden Layer 1: 50 neurons	Hidden Layer 1: 80 neurons
With Adagrad	MSE Loss