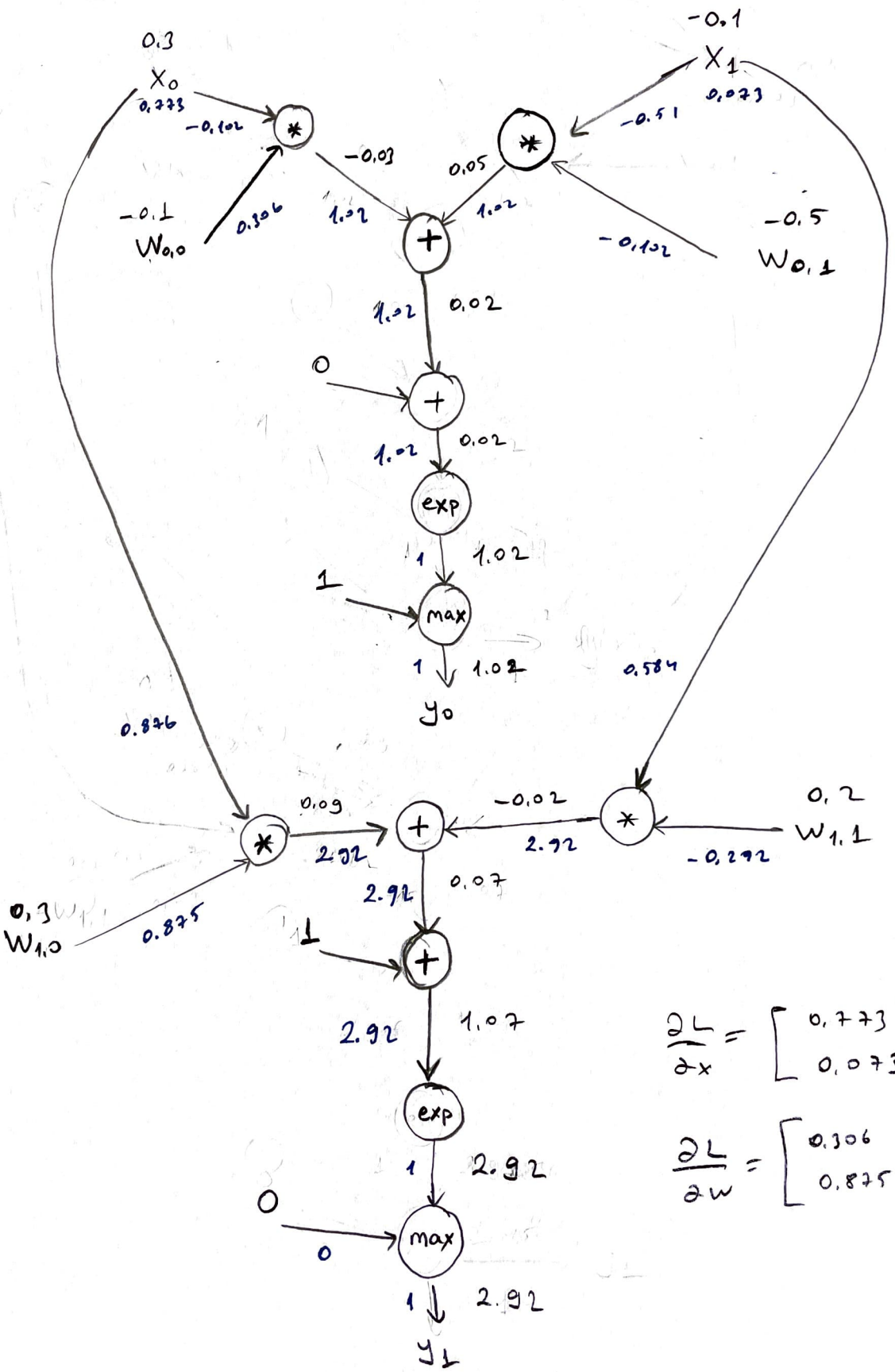
Barış Deniz Sağlam
155841O

Q1)

The gradients are written in blue.



$$\frac{\partial L}{\partial x} = \begin{bmatrix} 0.773 \\ 0.073 \end{bmatrix}$$

$$\frac{\partial L}{\partial w} = \begin{bmatrix} 0.306 & -0.102 \\ 0.875 & -0.292 \end{bmatrix}$$

Boris Deniz JAÇLAM
155841

Q2)

a) $y = ReLU(W^T . x + b)$

for given data points $z = W^T . x + b = -0.04$

since $ReLU(x) = max(0, x)$

it doesn't pass the output of neuron. Hence the gradients don't flow backwards. Mathematically

$$\frac{\partial\, ReLU(x)}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$$

$z = W^T . x + b$

$y = ReLU(z)$

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial W}$$

$$\left.\frac{\partial y}{\partial z}\right|_{\text{at given } x} = 0 \quad \text{hence} \quad \frac{\partial L}{\partial W} = 0$$

b) Yes. If in the next iterations there is such an x that

$$W^T . x + b > 0$$

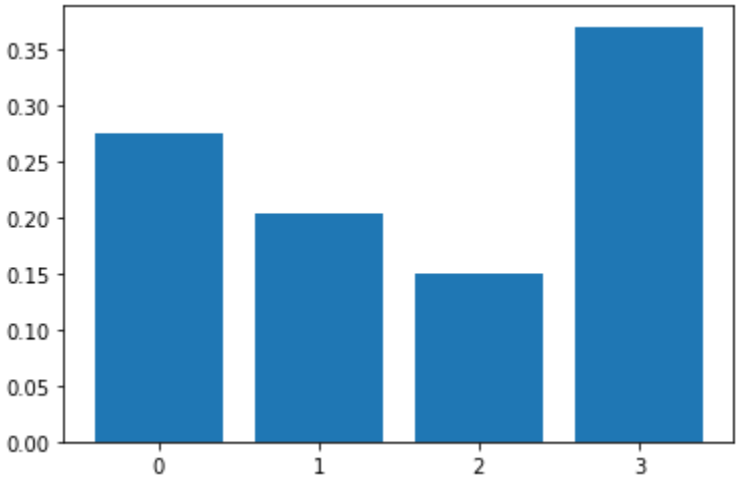
then $$\frac{\partial L}{\partial W} \neq 0$$

hence, the neuron learns.

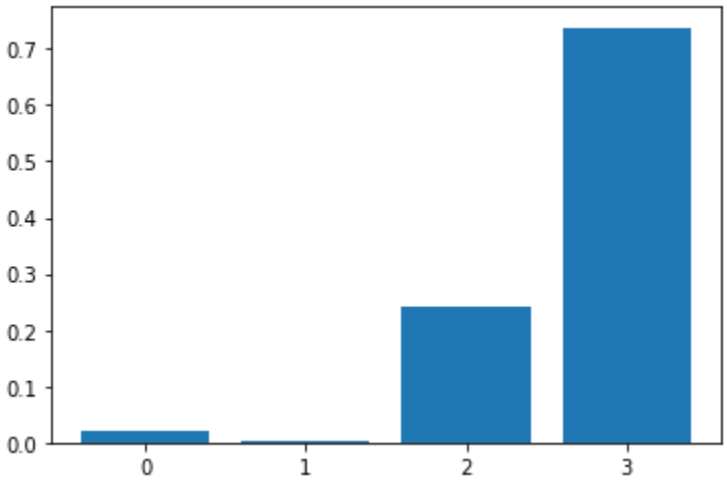
Barış Deniz Sağlam
1558410
01.12.2021

Question 3

a) [0.27476388, 0.2035501, 0.1507936, 0.37089244]

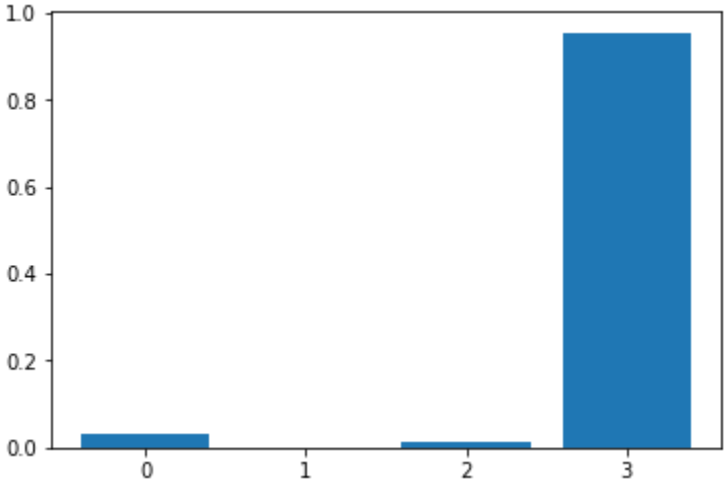


b) [0.01917612, 0.00230174, 0.24072742, 0.73779476]

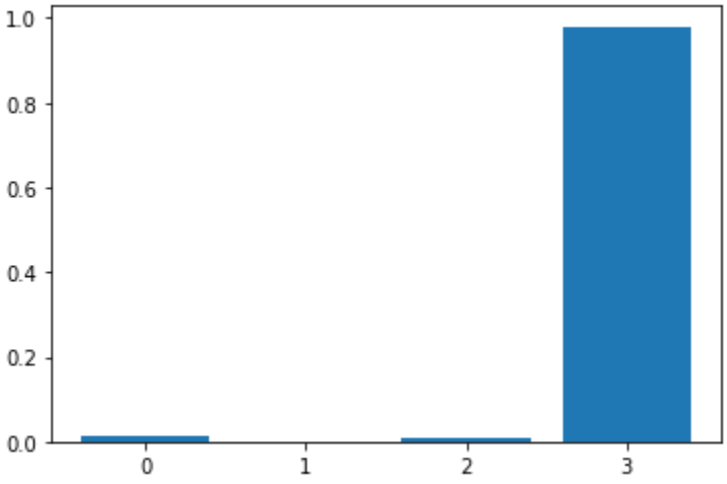


c) [3.2188911e-02, 1.9624767e-04, 1.2700258e-02, 9.5491463e-01]

Barış Deniz Sağlam
1558410
01.12.2021



d) [1.3429911e-02, 1.6488365e-04, 6.4719976e-03, 9.7993314e-01]



Barış Deniz Sağlam
1558410

Question 4

no	name	type	parameters	activations	learnables
L0	data	input		3001 x 33 x 1	
L1	conv1_1	Conv + ReLU	filters: 64 size: 200 x 1 x 1 padding: [3 3 ; 0 0] stride: [1 1]	2808 x 33 x 64	Weights: 200 x 1 x 1 x 64 Bias: 1 x 1 x 64
L2	conv1_2	Conv + ReLU	filters: 64 size: 3 x 1 x 64 padding: [1 1 ; 0 0] stride: [1 1]	2808 x 33 x 64	Weights: 3 x 1 x 64 x 64 Bias: 1 x 1 x 64
L2	pool1	Max-Pooling	size: 6 x 1 padding: [0 0 ; 0 0] stride: [6 1]	468 x 33 x 64	Weights: 0 Bias: 0
L3	conv2_1	Conv + ReLU	filters: 128 size: 3 x 1 x 64 padding: [1 1 ;0 0] stride: [1 1]	468 x 33 x 128	Weights: 3 x 1 x 64 x 128 Bias: 1 x 1 x 128
L4	conv2_2	Conv + ReLU	filters: 128 size: 3 x 1 x 128 padding: [1 1 ; 0 0] stride: [1 1]	468 x 33 x 128	Weights: 3 x 1 x 128 x 128 Bias: 1 x 1 x 128
L4	pooling2	Max-Pooling	size: 6 x 1 padding: [0 0 0 0] stride: [6 1]	78 x 33 x 128	Weights: 0 Bias: 0

Barış Deniz Sağlam
1558410

L5	conv3_1	Conv + ReLU	filters: 128 size: 3 x 1 x 128 padding: [1 1 ; 0 0] stride: [1 1]	78 x 33 x 128	Weights: 3 x 1 x 128 x 128 Bias: 1 x 1 x 128
L6	conv3_2	Conv + ReLU	filters: 128 size: 3 x 2 x 128 padding: [1 1 ; 1 1] stride: [1 3]	78 x 12 x 128	Weights: 3 x 2 x 128 x 128 Bias: 1 x 1 x 128
L6	pool3	Max-Pooling	size: 6 x 1 padding: [0 0 0 0] stride: [6 1]	13 x 12 x 128	Weights: 0 Bias: 0
L7	fc7	FC	nodes: 256	256	Weights: 256 x 19968 Bias: 256
L8	fc8	FC	nodes: 256	256	Weights: 256 x 256 Bias: 256
L9	fc9	FC	nodes: 10	10	Weights: 10 x 256 Bias: 10
L10	prob	Soft-max	nodes: 10	10	Weights: 0 Bias: 0

Barış Deniz Sağlam
1558410
01.12.2021

Question 5

a) Each epoch takes 155 iterations, then there must be roughly 155 * 32 = **4960** samples in the training set.

b) The model is learning as both validation loss and validation RMSE are decreasing. Hence, it's not overfitting. However, it is learning slowly, as the validation loss decreases by a small amount with each iteration. I would experiment with larger learning rates and choose the largest learning rate that doesn't diverge the training.