Assignment#1

Group Members

Muhammad Hamza jamil	**2572890**	**s8mujami@stud.uni-saarland.de**
Hacane Hechehouche	**2571617**	**S8hahech@stud.uni-Saarland.de**

Exercise 7.1

a) we have to chose the line that provide the biggest distance from the support vectors (the closest points to the line). so that we pick (3) because it provide the biggest distance from the support vectors.

b) $w = \sum_i \alpha_i s_i$ for support vectors. that are.

$s_1 = \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix}$ $s_2 = \begin{bmatrix} 3 \\ 3,5 \end{bmatrix}$ $s_3 = \begin{bmatrix} 3 \\ 1,5 \end{bmatrix}$ $s_4 = \begin{bmatrix} 3,5 \\ 2,5 \end{bmatrix}$

$$w = \sum \alpha_i s_i = 2,48 \cdot \begin{bmatrix} 2,5 \\ 2,5 \end{bmatrix} + 0,02 \begin{bmatrix} 3 \\ 3,5 \end{bmatrix} + (-0,98) \begin{bmatrix} 3 \\ 1,5 \end{bmatrix} + (-1,52) \begin{bmatrix} 3,5 \\ 2,5 \end{bmatrix}$$

$$w = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

$\boxed{P_1}$

$$y = w^T x + b$$

$$\begin{bmatrix} -2 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1,5 \end{bmatrix} + 3,15 = -3$$

sign is negative $\rightarrow$ lower class.

$\boxed{P_2}$ $y = wx + b$

$$= \begin{bmatrix} -2 \\ 1 \end{bmatrix}^T \begin{bmatrix} 3 \\ 2{,}5 \end{bmatrix} + 3{,}5 \quad = -6 + 2{,}5 + 3{,}5 = 0.$$

this points lies on the line.

$\boxed{C}$

- Maximum margin: it's called this way because its basically maximizing of the margin between the hyperplane and the support vectors. the hypeplane with maximum margin is what we are looking for.

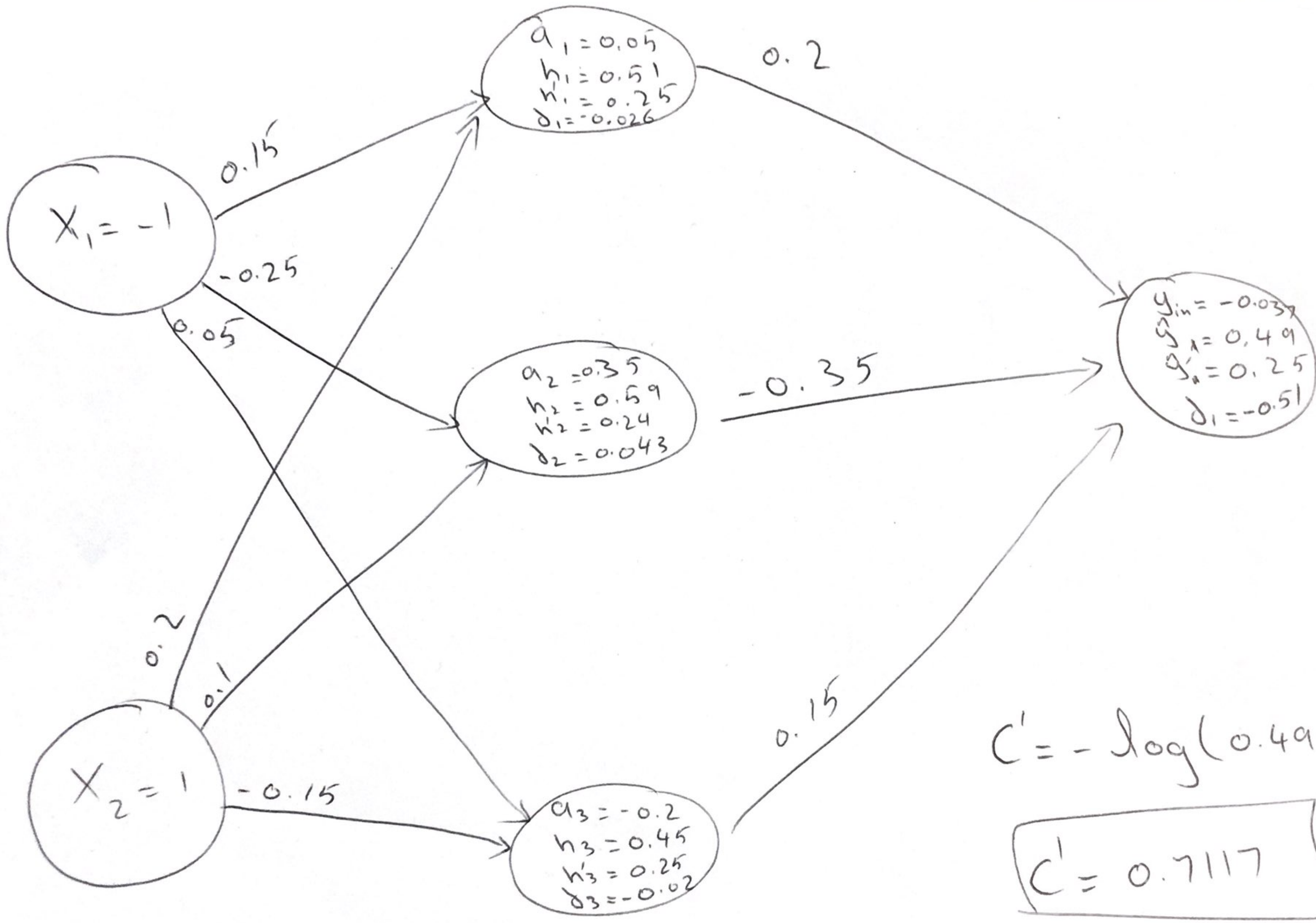
- "support vectors" are the critical points for the hyperplane. they are the points that support the hyperplane from all sides. and they decide the location of it. changing their location changes the hyperplane.

$\boxed{D}$ we have 2 options:

① we can use non-linear hyperplane that separate the data.

② we can use a kernel that add a new dimension to the data to make it linearly separable in the new dimension like dot product or distance from the center of data like the example in the lecture.

9)



$$C' = -\log(0.49)$$

$$\boxed{C' = 0.7117}$$

b) Back Propogation

$$\frac{\partial L}{\partial \hat{y}_i} = \frac{1}{n} \sum_{i=1}^{n} -\frac{y_i}{\hat{y}_i} + \frac{(1-y_i)}{(1-\hat{y}_i)}$$

$$= -\frac{1}{0.49} + 0$$

$$\boxed{\frac{\partial L}{\partial y_1} = -2.04}$$

For output Layer

$$\delta_1 = \frac{\partial L}{\partial \hat{y}_1} \times \frac{\partial}{\partial a_j} h(a_j)$$

$$\delta_1 = -2.04 \times 0.25$$

$$\boxed{\delta_1 = -0.51}$$

For Weights of output layer

$$\frac{\partial E}{\partial w_1} = \delta_1 \times h_1^1$$

$$= -0.51 \times 0.51$$

$$\boxed{\frac{\partial E}{\partial w_1} = -0.26}$$

$$\frac{\partial E}{\partial w_2} = \delta_1 \times h_2^1$$

$$= -0.51 \times 0.59$$

$$\boxed{\frac{\partial E}{\partial w_2} = -0.30}$$

②

$$\frac{\partial E}{\partial w_3} = \delta_1 \times h_3^1$$

$$= -0.53 \times 0.45$$

$$\boxed{\frac{\partial E}{\partial w_3} = -0.23}$$

Gradient Desent / UPdated weights

$w_1 = 0.2 + (0.1 \times -0.26)$

$\boxed{w_1 = -0.174}$

$w_2 = 0.35 + (0.1 \times -0.30)$

$\boxed{w_2 = -0.38}$

$w_3 = 0.15 + (0.1 \times -0.23)$

$\boxed{w_3 = 0.127}$

For Hidden Layer

$\delta_1 = (-0.53 \times 0.2 \times 0.25)$

$\boxed{\delta_1 = -0.106} \times 0.25 \Rightarrow \boxed{-0.026}$

$\delta_2 = -0.53 \times -0.35 \times 0.24$

$\boxed{\delta_2 = 0.18} \times 0.24 \Rightarrow \boxed{0.043}$

$\delta_3 = -0.53 \times 0.15 \times 0.25$

$\boxed{\delta_3 = -0.08} \times 0.25 \Rightarrow \boxed{-0.02}$

For Weights of hidden Layer

$\boxed{w_1'}$

$$\frac{\partial E}{\partial w_1'} = -0.026 \times -1 \Rightarrow \boxed{0.026}$$

$$w_1'' = w_1' + \varepsilon \frac{\partial E}{\partial w_1'}$$

$$w_1'' = 0.15 + 0.1 \times 0.026$$

$$\boxed{w_1' = 0.176}$$

For w_2'

$$\frac{\partial E}{\partial w_2'} = 0.043 \times -1 \Rightarrow -0.043$$

$$w_2'' = -0.25 + 0.1 \times -0.043$$

$$\boxed{w_2'' = -0.2543}$$

For w_3'

$$\frac{\partial E}{\partial w_3'} = -0.02 \times -1 \Rightarrow 0.02$$

$$w_3'' = 0.05 + 0.1 \times 0.02$$

$$\boxed{w_3'' = 0.052}$$

Gradient ~~Desent for hidden~~

For w^2_1

$$\frac{\partial E}{\partial w^2_1} = -0.026 \times 1 \Rightarrow -0.026$$

$$w'^2_1 = w^2_1 + \varepsilon \frac{\partial E}{\partial w^2_1}$$

$$= +0.2 + 0.1 \times -0.026$$

$$\boxed{w'^2_1 = 0.197}$$

For w^2_2

$$\frac{\partial E}{\partial w^2_2} = 0.043$$

$$w'^2_2 = w^2_2 + \varepsilon \frac{\partial E}{\partial w^2_2}$$

$$w'^2_2 = 0.1 + 0.1 \times 0.043$$

$$w'^2_2 = 0.1043$$

For w^2_3

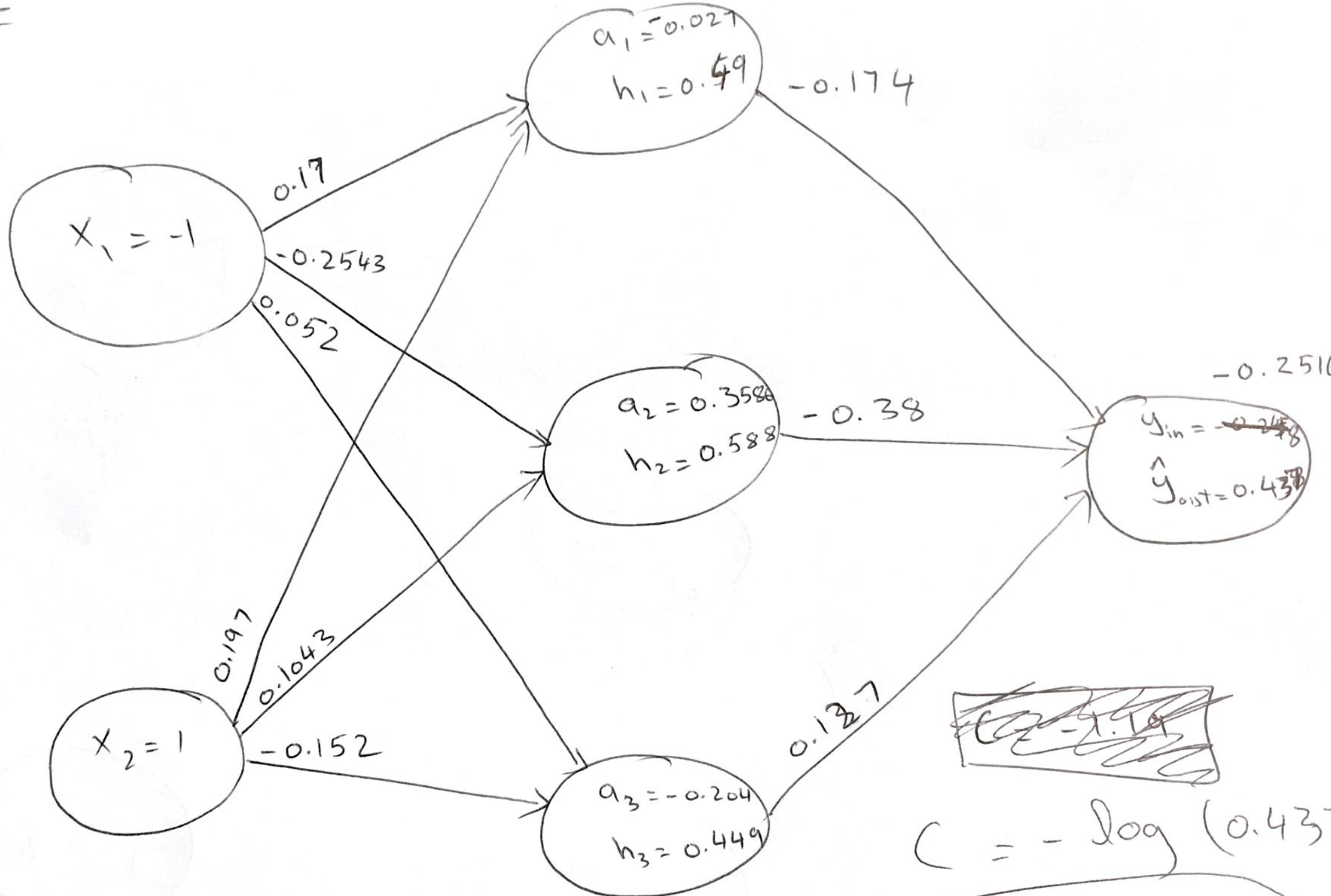
$$\frac{\partial E}{\partial w^2_3} = -0.02$$

$$w'^2_3 = -0.15 + 0.1 \times -0.02$$

$$\boxed{w'^2_3 = -0.152}$$

Step 3

5



$C = -\log(0.437)$

$\boxed{C = -0.8278}$