Ilir Gusija

20158525

CISC/CMPE 452/COGS 400
Assignment 1
Theoretical Part

1. **Percptron:** - Consider the classification problem defined below.
$\mathbf{p}_1 = ([-1,1]^t, 1)$, $\mathbf{p}_2 = ([-1,-1]^t, 1)$, $\mathbf{p}_3 = ([0,0]^t, 0)$, and $\mathbf{p}_4 = ([1,0]^t, 0)$.
   (a) Design a single-neuron perceptron to solve this problem. Design the network graphically, by choosing weight vectors that are orthogonal to the decision boundaries.
   (b) Test your solution with all four input vectors.
   (c) Classify the following input vectors with your solution $\mathbf{p}_5 = ([-2,0]^t)$, $\mathbf{p}_6 = ([1,1]^t)$, $\mathbf{p}_7 = ([0,1]^t)$, and $\mathbf{p}_8 = ([-1,-2]^t)$.
   (d) Which of the vectors in part (3) will always be classified the same way, regardless of the solution? Which may vary depending on the solution? Why?

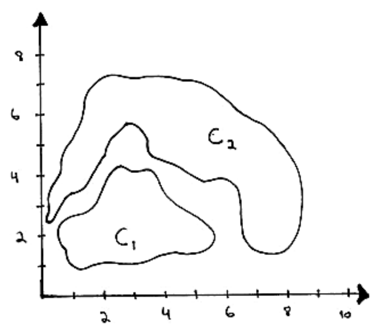
2. **Multilayer Percptron** - You are presented with the following input set:
$\mathbf{x}_1 = [3,1,1]^t$, $\mathbf{x}_2 = [4,0,1]^t$, $\mathbf{x}_3 = [4,-1,1]^t$, $\mathbf{x}_4 = [5,2,1]^t$,$\mathbf{x}_5 = [5,3,1]^t$, $\mathbf{x}_6 = [3,3,1]^t$, $\mathbf{x}_7 = [2,0,1]^t$, and $\mathbf{x}_8 = [1,1,1]^t$. A neural network with two discrete bipolar perceptrons in the hidden layer and a single discrete bipolar output perceptron needs to classify the presented inputs in either of two classes, C_1 and C_2 such that $x_1, x_2, x_3 \in C_1$, with the remaining inputs belonging to C_2.

   i– Check whether the weights $\mathbf{w}_1 = [2,1]^T, \theta_1 = 5, \quad \text{and} \quad \mathbf{w}_2 = [0,1]^T, \theta_2 = -2$ would provide the linear separation of patterns as required.
   ii– Repeat part **i** for the weights $\mathbf{w}_1 = [0,-1]^T, \theta_1 = 1.5, \quad \text{and} \quad \mathbf{w}_2 = [1,0]^T, \theta_2 = -2.5$
   iii– Complete the design of the classifier by using the results from either part **i** or **ii** and compute the weights of the single perceptron at the output.

3. **Constructing a Network** - Construct a multilayer perceptron which will be able to separate the two classes shown in the figure. Use two neurons in the output layer and find a suitable number of hidden layer neurons. Two output of the network should be $[1,0]^t$ if the input belongs to class C_1 and $[0,1]^t$ if the input belongs to class C_2. Use the activation function

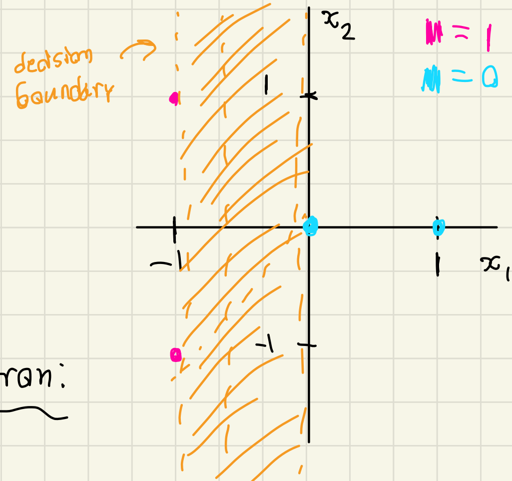
$$\sigma(net) = \begin{cases} 1 & \text{if } net > 0; \\ 0 & \text{if } net \leq 0. \end{cases}$$

and determine the weight by hand. What is the minimum amount of neurons in the hidden layer required for a perfect separation of the classes?



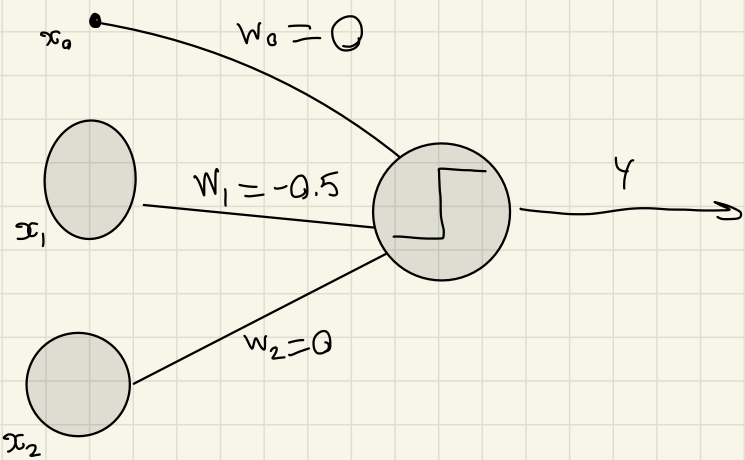
1. Given

$$\begin{array}{ll} & \quad x \qquad\quad | \; y \\ p_1 = & ([-1,1]^t,\ 1) \\ p_2 = & ([-1,-1]^t,\ 1) \\ p_3 = & ([0,0]^t,\ 0) \\ p_4 = & ([1,0]^t,\ 0) \end{array}$$



(a) Design single-neuron perceptron:

First we observe that the given data is linearly separable.



b) model $= f([x_1, x_2]^t) = \begin{cases} 1 & \text{if } -0.5 \cdot x_1 > 0 \\ 0 & \text{if } -0.5 \cdot x_1 \leq 0 \end{cases}$

$\Rightarrow f(p_1) = f([-1,1]^t) = 1$ since $-0.5 \cdot x_1 = -0.5 \cdot -1 = 0.5 > 0$

$f(p_2) = f([-1,-1]^t) = 1$ since $-0.5 \cdot x_1 = -0.5 \cdot -1 = 0.5 > 0$

$f(p_3) = f([0,0]^t) = 0$ since $-0.5 \cdot x_1 = -0.5 \cdot 0 = 0 \leq 0$

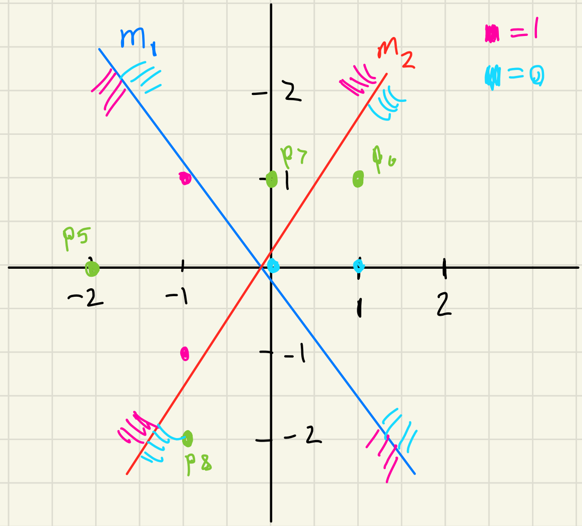
$f(p_4) = f([1,0]^t) = 0$ since $-0.5 \cdot x_1 = -0.5 \cdot 1 = -0.5 \leq 0$

c) $f(p_5) = f([-2,0]^t) = 1$

$f(p_6) = f([1,1]^t) = 0$

$f(p_7) = f([0,1]^t) = 0$

$f(p_8) = f([-1,-2]^t) = 1$



(d) Points p_5 and p_6 will always be classified as 1 and 0 respectively that is because if otherwise the data is no longer linearly separable.

For points p_7 & p_8 you could use decision boundaries such as m_1 and m_2 (which are displayed in the graph above) for which the use of m_1 has $f(p_7) = 0$ & $f(p_8) = 1$ and

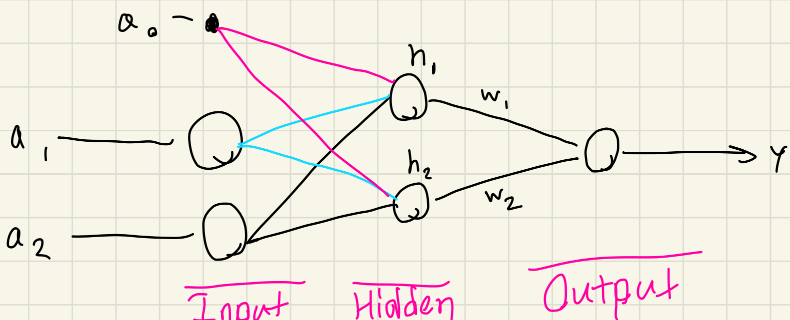
m_2 has $f(p_7) = 1$ & $f(p_8) = 0$

$\square$

2.

$$C_1 \begin{cases} x_1 = [3,1,1]^T \\ x_2 = [4,0,1]^T \\ x_3 = [4,-1,1]^T \end{cases} \qquad \left. \begin{matrix} x_4 = [5,2,1]^T \\ x_5 = [5,3\ 1]^T \\ x_6 = [3,3,1]^T \\ x_7 = [2,0,1]^T \\ x_8 = [1,1,1]^T \end{matrix} \right\} C_2$$

$$x_i = [a_2, a_1, a_0]^T$$



i) $W_1 = [2,1]^T \quad \theta_1 = 5 \qquad W_2 [0,1]^T \quad \theta_2 = -2$

x_1: $\begin{bmatrix}3\\1\end{bmatrix} \cdot \begin{bmatrix}2\\1\end{bmatrix} + 5 = 12 \geq 0 = 1$ $\qquad \begin{bmatrix}3\\1\end{bmatrix} \cdot \begin{bmatrix}0\\1\end{bmatrix} - 2 = -1 \leq 0 = -1$

x_2: $\begin{bmatrix}4\\0\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 13 \geq 0 = 1$ $\qquad \begin{bmatrix}4\\0\end{bmatrix} \cdot \begin{bmatrix}0\\1\end{bmatrix} - 2 = -2 \leq 0 = -1$

x_3: $\begin{bmatrix}4\\-1\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 12 \geq 0 = 1$ $\qquad \begin{bmatrix}4\\-1\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = -3 \leq 0 = -1$

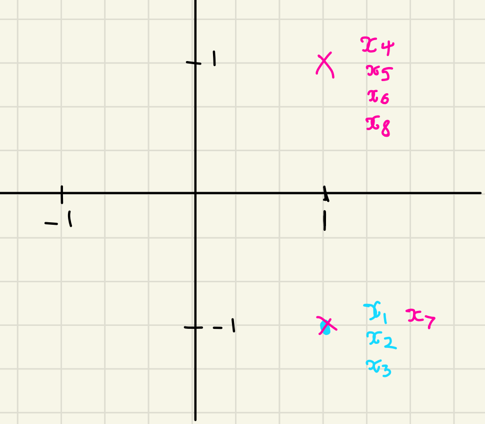
x_4: $\begin{bmatrix}5\\2\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 17 \geq 0 = 1$ $\qquad \begin{bmatrix}5\\2\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = 0 \geq 0 = 1$

x_5: $\begin{bmatrix}5\\3\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 18 \geq 0 = 1$ $\qquad \begin{bmatrix}5\\3\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = 1 \geq 0 = 1$

x_6: $\begin{bmatrix}3\\3\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 14 \geq 0 = 1$ $\qquad \begin{bmatrix}3\\3\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = 1 \geq 0 = 1$

x_7: $\begin{bmatrix}2\\0\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 9 \geq 0 = 1$ $\qquad \begin{bmatrix}2\\0\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = -2 \leq 0 = -1$

x_8: $\begin{bmatrix}1\\1\end{bmatrix} \begin{bmatrix}2\\1\end{bmatrix} + 5 = 8 \geq 0 = 1$ $\qquad \begin{bmatrix}1\\1\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} - 2 = 1 \geq 0 = 1$



As we can see, these weights do not provide the required linear separation of patterns required.

$\square$

ii) $W_1 = [0, -1]^T \quad \theta_1 = 1.5 \qquad w_2 = [1, 0]^T \quad \theta_2 = -2.5$

x_1: $\begin{bmatrix}3\\1\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = 0.5 \geq 0 = 1 \qquad \begin{bmatrix}3\\1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 0.5 \geq 0 = 1.$

x_2: $\begin{bmatrix}4\\0\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = 1.5 \geq 0 = 1 \qquad \begin{bmatrix}4\\0\end{bmatrix}\cdot\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 1.5 \geq 0 = 1$

x_3: $\begin{bmatrix}4\\-1\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = 2.5 \geq 0 = 1 \qquad \begin{bmatrix}4\\-1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 1.5 \geq 0 = 1$

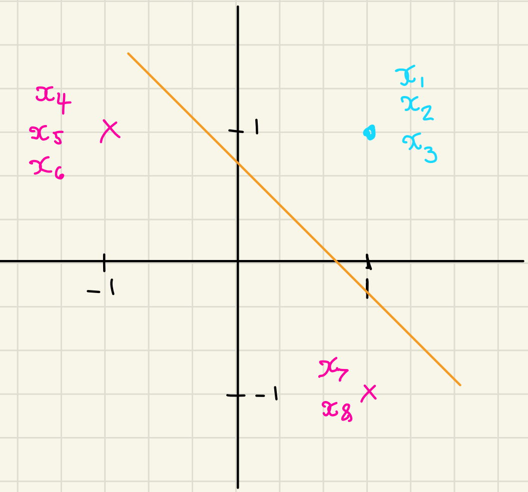
x_4: $\begin{bmatrix}5\\2\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = -0.5 \leq 0 = -1 \qquad \begin{bmatrix}5\\2\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 2.5 \geq 0 = 1$

x_5: $\begin{bmatrix}5\\3\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = -1.5 \leq 0 = -1 \qquad \begin{bmatrix}5\\3\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 2.5 \geq 0 = 1$

x_6: $\begin{bmatrix}3\\3\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = -1.5 \leq 0 = -1 \qquad \begin{bmatrix}3\\3\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = 0.5 \geq 0 = 1$

x_7: $\begin{bmatrix}2\\0\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = 1.5 \geq 0 = 1 \qquad \begin{bmatrix}2\\0\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = -0.5 \leq 0 = -1$

x_8: $\begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}+1.5 = 0.5 \geq 0 = 1 \qquad \begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}-2.5 = -1.5 \leq 0 = -1$



These weights do provide the linear separation of patterns that's required

$\square$

iii) Let $i_1, i_2 \in \{-1, 1\}$, $f: \{-1,1\} \longrightarrow \{-1,1\}$ s.t.

$$f(i_1, i_2) = \begin{cases} 1, & i_1 = i_2 = 1 \\ -1, & \text{otherwise} \end{cases}$$

Then using the weights from ii) where

$$y_{h_1}(\vec{x}) = \begin{bmatrix} 0 \\ -1 \\ 1.5 \end{bmatrix} \vec{x} \qquad , \; y_{h_2}(\vec{x}) = \begin{bmatrix} 1 \\ 0 \\ -2.5 \end{bmatrix} \vec{x}$$

denote the outputs of the hidden layer & the inputs to the output node. We have for each input vector;

$\vec{x} = \vec{x}_1 \qquad f(y_{h_1}(\vec{x}_1), y_{h_2}(\vec{x}_1)) = f(1,1) = 1$

$\vec{x}_2 \qquad " \qquad f(1,1) = 1$

$\vec{x}_3 \qquad " \qquad f(1,1) = 1$

$\vec{x}_4 \qquad " \qquad f(-1,1) = -1$

$\vec{x}_5 \qquad " \qquad " = -1$

$\vec{x}_6 \qquad " \qquad " = -1$

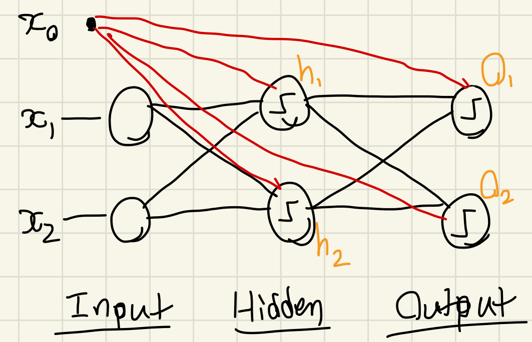
$\vec{x}_7 \qquad " \qquad f(1,-1) = -1$

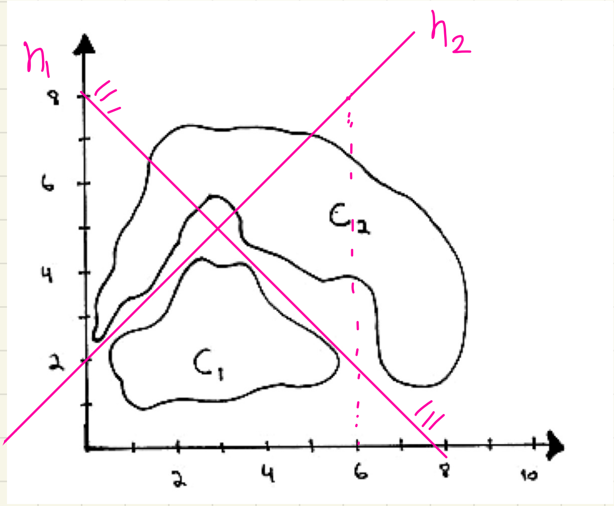
$\vec{x}_8 \qquad " \qquad " = -1$

So if we take an output of 1 to mean a class of C_1 & -1 to mean a class of C_2 we have that the network correctly classifies all inputs.

$\square$

3.





h_1: $x_2 = -x_1 + 8$ $\Rightarrow$ $w^{21}_{1i} = [1, 1]$ $\theta^{21}_{1i} = -8$

h_2: $x_2 = x_1 + 2$ $w^{21}_{2i} = [1, -1]$ $\theta^{21}_{2i} = -2$

O_1: $x_2 = -x_1 + 1/2$ $w^{32}_{1h} = [-1, -1]$ $\theta^{32}_{1h} = 1/2$

O_2: $-x_2 = x_1 - 1/2$ $w^{32}_{2h} = [1, 1]$ $\theta^{32}_{2h} = -1/2$

$$h_1 = \sigma(x_1 + x_2 - 8) \qquad O_1 = \sigma(h_1 + h_2 - 1/2)$$
$$h_2 = \sigma(x_1 - x_2 - 2) \qquad O_2 = \sigma(-(h_1 + h_2 - 1/2))$$

We find that the two classes are separable when using two straight lines to form a decision boundary hence the minimum amount of hidden nodes is <u>two</u>. The weights for each node are shown above. $\square$