hw1

Hongyi Guo 516030910306

March 7, 2019

1

1.1

$\mathbf{w} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, $b = -1/2$

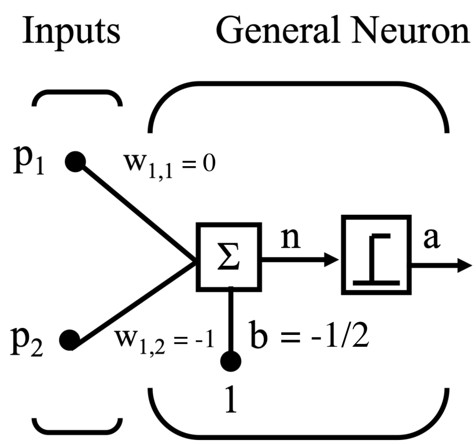

Figure 1: Designed perceptron

1.2

$$a_1 = \text{hardlim}(\mathbf{w}^T\mathbf{p_1} + b) = \text{hardlim}\left(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} - 1/2\right) = \text{hardlim}(1/2) = 1 = t_1$$

$$a_2 = \text{hardlim}(\mathbf{w}^T\mathbf{p_2} + b) = \text{hardlim}\left(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} -1 \\ -1 \end{bmatrix} - 1/2\right) = \text{hardlim}(1/2) = 1 = t_2$$

$$a_3 = \text{hardlim}(\mathbf{w}^T\mathbf{p_3} + b) = \text{hardlim}\left(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 1/2\right) = \text{hardlim}(-1/2) = 0 = t_3$$

$$a_4 = \text{hardlim}(\mathbf{w}^T\mathbf{p_4} + b) = \text{hardlim}\left(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1/2\right) = \text{hardlim}(-1/2) = 0 = t_4$$

1.3

$$a_5 = \text{hardlim}(\mathbf{w}^T\mathbf{p_5} + b) = \text{hardlim}(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} -2 \\ 0 \end{bmatrix} - 1/2) = \text{hardlim}(-1/2) = 0$$

$$a_6 = \text{hardlim}(\mathbf{w}^T\mathbf{p_6} + b) = \text{hardlim}(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1/2) = \text{hardlim}(-3/2) = 0$$

$$a_7 = \text{hardlim}(\mathbf{w}^T\mathbf{p_7} + b) = \text{hardlim}(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 1/2) = \text{hardlim}(-3/2) = 0$$

$a_8 = \text{hardlim}(\mathbf{w}^T\mathbf{p_8} + b) = \text{hardlim}(\begin{bmatrix} 0 & -1 \end{bmatrix} \times \begin{bmatrix} -1 \\ -2 \end{bmatrix} - 1/2) = \text{hardlim}(3/2) = 1$

1.4

Classification of P_6, P_7, P_8 is irrelevant to $\mathbf{w}$ and b, but P_5 is relevant.
Use P_7 as an example. We prove it is irrelevant to $\mathbf{w}$ and b via proving it cannot belong to label 1.
Now assuming P_7 belongs to 1. We draw convex hulls of label 0 and label 1 as in Fig.2. It's easy to see the two convex hulls are overlapping. So our data are not linear seperatable. Thus, P_7 cannot belong to label 1. The proof of P_6 and P_8 is similar to P_7. For P_5, whichever it belongs to, 0 or 1, the two convex hulls are not overlapping.

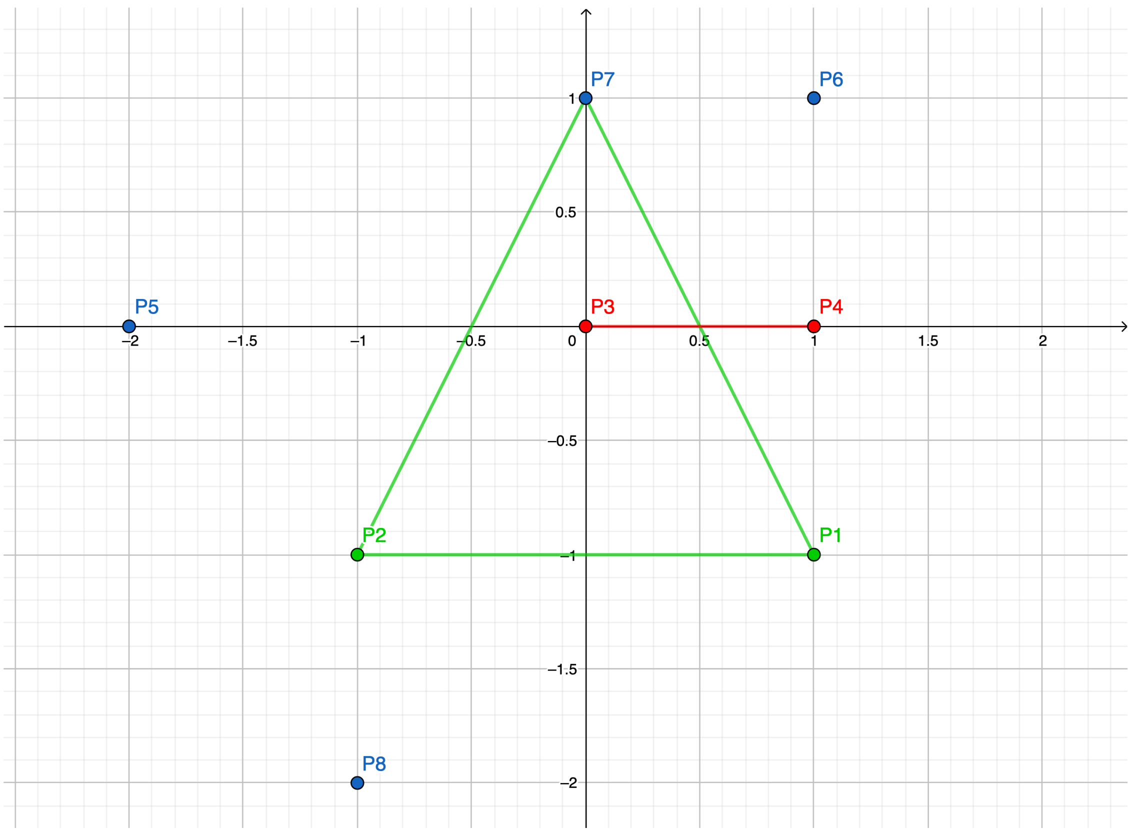

Figure 2: P7 cannot be labeled 1

1.5

1. *Present P1 to the network:*

   $a_1 = \text{hardlim}(\mathbf{w}^T\mathbf{p_1} + b) = \text{hardlim}(\begin{bmatrix} 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 0) = \text{hardlim}(0) = 1 = t_1$

2. *Present P2 to the network:*

   $a_2 = \text{hardlim}(\mathbf{w}^T\mathbf{p_2} + b) = \text{hardlim}(\begin{bmatrix} 0 & 0 \end{bmatrix} \times \begin{bmatrix} -1 \\ -1 \end{bmatrix} + 0) = \text{hardlim}(0) = 1 = t2$

3. *Present P3 to the network:*

   $a_3 = \text{hardlim}(\mathbf{w}^T\mathbf{p_3} + b) = \text{hardlim}(\begin{bmatrix} 0 & 0 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \end{bmatrix} + 0) = \text{hardlim}(0) = 1$

$e_3 = t_3 - a_3 = 0 - 1 = -1$

$\mathbf{w} = \mathbf{w} + e_3\mathbf{p_3} = \begin{bmatrix}0\\0\end{bmatrix} - \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$

$b = b + e_3 = -1$

4. *Present P4 to the network:*

   $a_4 = \text{hardlim}(\mathbf{w}^T\mathbf{p_4} + b) = \text{hardlim}(\begin{bmatrix}0 & 0\end{bmatrix} \times \begin{bmatrix}1\\0\end{bmatrix} - 1) = \text{hardlim}(-1) = 0 = t_4$

5. *Present P1 to the network:*

   $a_1 = \text{hardlim}(\mathbf{w}^T\mathbf{p_1} + b) = \text{hardlim}(\begin{bmatrix}0 & 0\end{bmatrix} \times \begin{bmatrix}1\\-1\end{bmatrix} - 1) = \text{hardlim}(-1) = 0$

   $e_1 = t_1 - a_1 = 1 - 0 = 1$

   $\mathbf{w} = \mathbf{w} + e_1 * \mathbf{p_2} = \begin{bmatrix}0\\0\end{bmatrix} + \begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix}$

   $b = b + e_1 = 0$

6. *Present P2 to the network:*

   $a_2 = \text{hardlim}(\mathbf{w}^T\mathbf{p_2} + b) = \text{hardlim}(\begin{bmatrix}1 & -1\end{bmatrix} \times \begin{bmatrix}-1\\-1\end{bmatrix} + 0) = \text{hardlim}(0) = 1 = t_2$

7. *Present P3 to the network:*

   $a_3 = \text{hardlim}(\mathbf{w}^T\mathbf{p_3} + b) = \text{hardlim}(\begin{bmatrix}1 & -1\end{bmatrix} \times \begin{bmatrix}0\\0\end{bmatrix} + 0) = \text{hardlim}(0) = 1$

   $e_3 = t_3 - a_3 = 0 - 1 = -1$

   $\mathbf{w} = \mathbf{w} + e_3 * \mathbf{p_3} = \begin{bmatrix}1\\-1\end{bmatrix} - \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}1\\-1\end{bmatrix}$

   $b = b + e_3 = 0 + (-1) = -1$

8. *Present P4 to the network:*

   $a_4 = \text{hardlim}(\mathbf{w}^T\mathbf{p_4} + b) = \text{hardlim}(\begin{bmatrix}1 & -1\end{bmatrix} \times \begin{bmatrix}1\\0\end{bmatrix} - 1) = \text{hardlim}(0) = 1$

   $e_4 = t_4 - a_4 = 0 - 1 = -1$

   $\mathbf{w} = \mathbf{w} + e_4 * \mathbf{p_4} = \begin{bmatrix}1\\-1\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\-1\end{bmatrix}$

   $b = b + e_4 = -1 + (-1) = -2$

9. *Present P1 to the network:*

   $a_1 = \text{hardlim}(\mathbf{w}^T\mathbf{p_1} + b) = \text{hardlim}(\begin{bmatrix}0 & -1\end{bmatrix} \times \begin{bmatrix}1\\-1\end{bmatrix} - 2) = \text{hardlim}(-1) = 0$

   $e_1 = t_1 - a_1 = 1 - 0 = 1$

   $\mathbf{w} = \mathbf{w} + e_1 * \mathbf{p_2} = \begin{bmatrix}0\\-1\end{bmatrix} + \begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}1\\-2\end{bmatrix}$

   $b = b + e_1 = -1$

10. *Present P2 to the network:*

    $a_2 = \text{hardlim}(\mathbf{w}^T\mathbf{p_2} + b) = \text{hardlim}(\begin{bmatrix}1 & -2\end{bmatrix} \times \begin{bmatrix}-1\\-1\end{bmatrix} - 1) = \text{hardlim}(0) = 1 = t_2$

11. *Present P3 to the network:*

    $a_3 = \text{hardlim}(\mathbf{w}^T\mathbf{p_3} + b) = \text{hardlim}(\begin{bmatrix}1 & -2\end{bmatrix} \times \begin{bmatrix}0\\0\end{bmatrix} - 1) = \text{hardlim}(-1) = 0 = t_3$

12. *Present P4 to the network:*

$$a_4 = \text{hardlim}(\mathbf{w}^T\mathbf{p_4} + b) = \text{hardlim}(\begin{bmatrix}1 & -2\end{bmatrix} \times \begin{bmatrix}1\\0\end{bmatrix} - 1) = \text{hardlim}(0) = 1$$

$$e_4 = t_4 - a_4 = 0 - 1 = -1$$

$$\mathbf{w} = \mathbf{w} + e_4 * \mathbf{p_4} = \begin{bmatrix}1\\-2\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}0\\-2\end{bmatrix}$$

$$b = b + e_4 = -1 + (-1) = -2$$

Test

$$a_1 = \text{hardlim}(\mathbf{w}^T\mathbf{p_1} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}1\\-1\end{bmatrix} - 2) = \text{hardlim}(0) = 1 = t_1$$

$$a_2 = \text{hardlim}(\mathbf{w}^T\mathbf{p_2} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}-1\\-1\end{bmatrix} - 2) = \text{hardlim}(0) = 1 = t_2$$

$$a_3 = \text{hardlim}(\mathbf{w}^T\mathbf{p_3} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}0\\0\end{bmatrix} - 2) = \text{hardlim}(-2) = 0 = t_3$$

$$a_4 = \text{hardlim}(\mathbf{w}^T\mathbf{p_4} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}1\\0\end{bmatrix} - 2) = \text{hardlim}(-2) = 0 = t_4$$

So, $\mathbf{w} = \begin{bmatrix}0\\-2\end{bmatrix}$, $b = -2$. Then,

$$a_5 = \text{hardlim}(\mathbf{w}^T\mathbf{p_5} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}-2\\0\end{bmatrix} - 2) = \text{hardlim}(-2) = 0$$

$$a_6 = \text{hardlim}(\mathbf{w}^T\mathbf{p_6} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}1\\1\end{bmatrix} - 2) = \text{hardlim}(-4) = 0$$

$$a_7 = \text{hardlim}(\mathbf{w}^T\mathbf{p_7} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}0\\1\end{bmatrix} - 2) = \text{hardlim}(-4) = 0$$

$$a_8 = \text{hardlim}(\mathbf{w}^T\mathbf{p_8} + b) = \text{hardlim}(\begin{bmatrix}0 & -2\end{bmatrix} \times \begin{bmatrix}-1\\-2\end{bmatrix} - 2) = \text{hardlim}(2) = 1$$

2

2.1

The generated data are as in Fig.3.
To make the results more clear, I trained my model for 20,000 epoches. And the accuracy curves with learning rate $0.1, 0.001, 0.0001$ are as Fig.4.

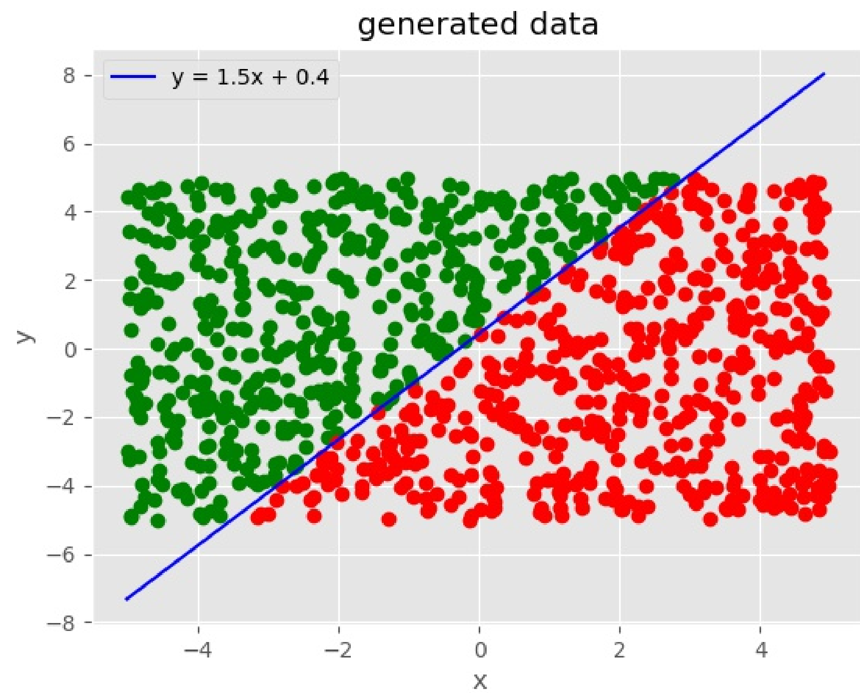

Figure 3: Generated data, where red dots are labeled -1 and green dots are labeled 1.

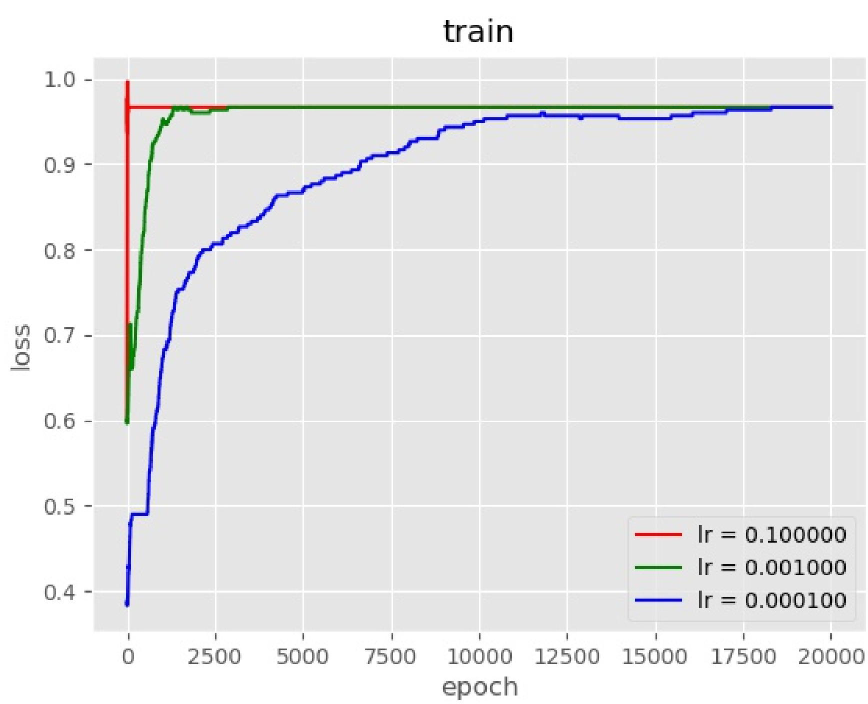

Figure 4: Result