Mid-Term Exam

2021 Fall

Name: TAN SIN HONG (202050817)

1. Depict the following training set in 3D space, and suggest whether linear separability is possible and why.

$$x_1 = \begin{pmatrix}0\\0\\0\end{pmatrix}, x_2 = \begin{pmatrix}0\\1\\1\end{pmatrix}, x_3 = \begin{pmatrix}1\\1\\1\end{pmatrix}, x_4 = \begin{pmatrix}1\\0\\1\end{pmatrix}, x_5 = \begin{pmatrix}0\\0\\1\end{pmatrix}, x_6 = \begin{pmatrix}0\\1\\0\end{pmatrix}$$

$$y_1 = 1, y_2 = 1, y_3 = 1, y_4 = -1, y_5 = -1, y_6 = -1$$

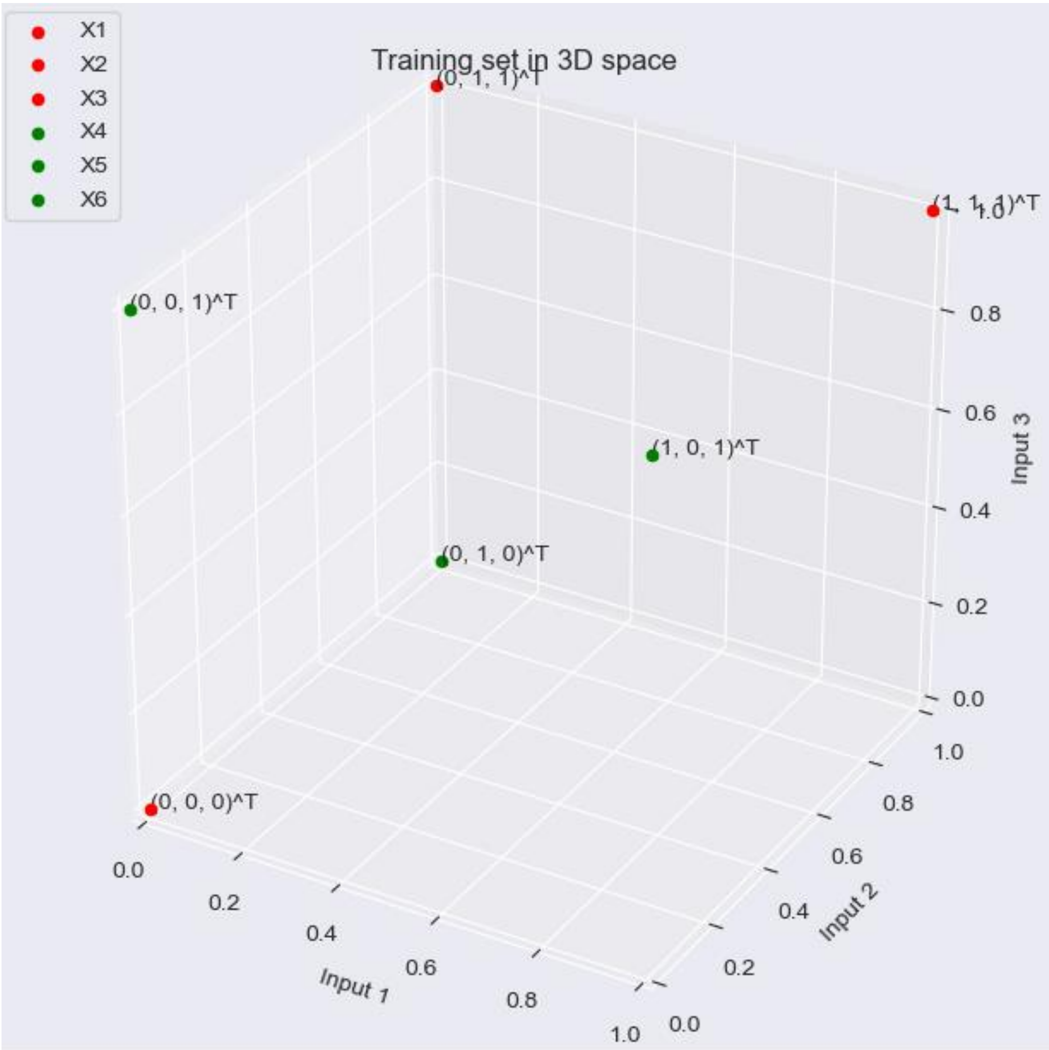
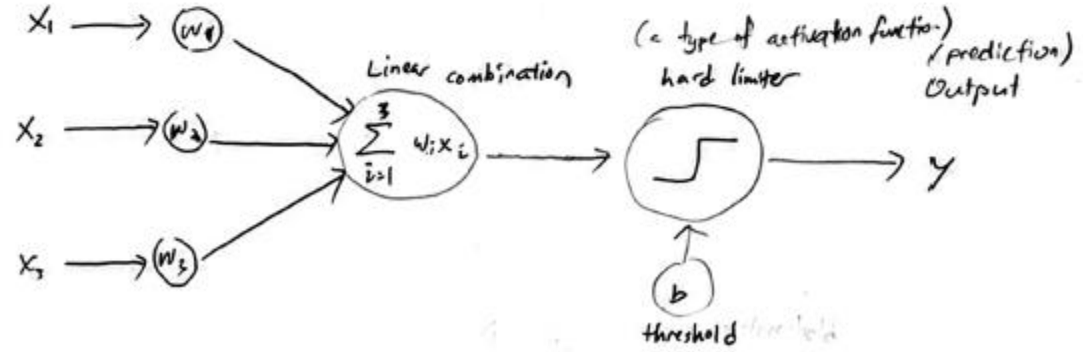


Figure 1 Training set in 3D space



Single layer perceptron with 3 input nodes.

- In the case of an elementary perceptron, the n-dimensional space is divided by a hyperplane into two decision regions. The hyperplane is defined by linearly separable function.

- The hyperplane equation is as defined below

$$w_1x_1 + w_2x_2 + w_3x_3 = b$$

- If this is generalized, the following equation can be the linear separable function

$$\sum_{i=0}^{3} w_ix_i - b = 0$$

In this case, $\quad w_1x_1 + w_2x_2 + w_3x_3 - b = 0$

If we express the hyperplane equation as matrix

$$(w_1, w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = b$$

- The normal vector perpendicular to the hyperplane is (w_1, w_2, w_3)
- and the distance from origin to the hyperplane is $\dfrac{|b|}{\|w\|_2}$

- If $(w_1, w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = b$, then the data is in hyperplane,

- If $(w_1 w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} > b$, then the data is located in the direction of upper part of the hyperplane.

- If $(w_1 w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} < b$, then the data is located in the direction of below part of the hyperplane.

- For a given dataset, if we can find the (w_1, w_2, w_3) that can satisfy the formula below, then we can say the given data is linear separable, if not, we can say the given data is non-linear-seperable.

$$Y \begin{cases} 1, & (w_1, w_2, w_3)\begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} \geq b \\ -1, & (w_1, w_2, w_3)\begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} \leq b \end{cases}$$

1. For $X_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

   $0 > b$

2. For $X_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$

   $w_2 + w_3 > b$

3. For $X_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

   $w_1 + w_2 + w_3 > b$

4. For $X_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

   $w_1 + w_3 \leq b$

5. For $X_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

   $w_3 \leq b$

6. For $X_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

   $w_2 \leq b$

∴ From the answer of (2) & (3)

$w_2 + w_3 > b$ — (2)

$w_1 + w_2 + w_3 > b$ — (3)

sub (2) into (3)

$w_1 > 0$

From the answer of (4) & (5)

$w_1 + w_3 \leq b$

$w_3 \leq b$

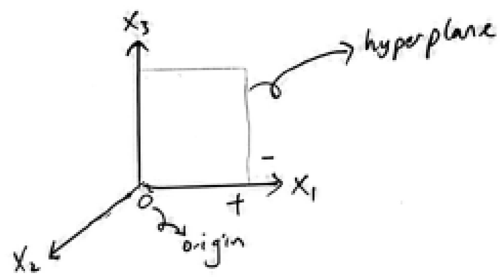
$w_1 < 0$

Since the equation of $w_1 > 0$ & $w_1 < 0$ is contradictory to each other, the given dataset cannot form linearly separable hyperplane. Therefore, we can say that the given data is non-linear-seperable.

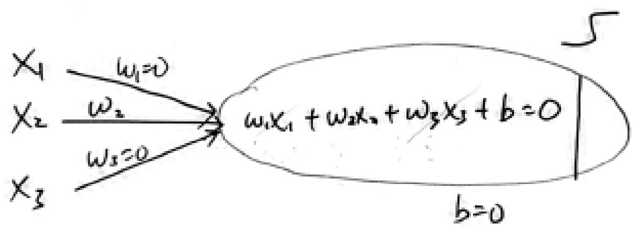
2. Solve the exercise problem 7 of Chapter 2 in textbook.

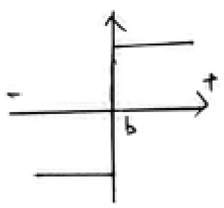
(The picture in the problem shows that the plane determined by the x1 and x3 axes. Present the corresponding perceptron as in Figure 2−4(a).)

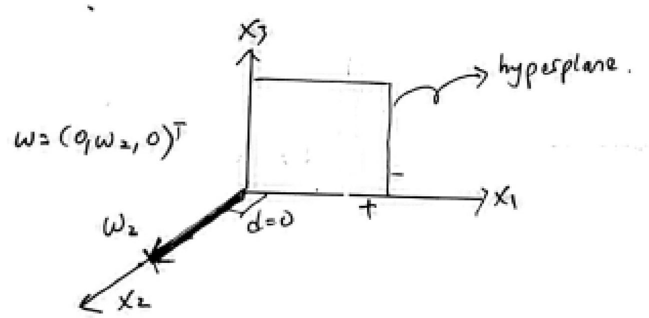
Q2.



- Since the hyperplane (in the figure) is parallel to x_1 & x_2, w_1, w_2 will be equal to 0.
- Also, since the distance of hyperplane from origin is 0, we can know the value of b will be 0.







The linear seperation between perceptron and hyperplane

3. Depict the following training set in 3D space. Present a perceptron that classifies this data with the least error rate.

$$x_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, x_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, x_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, x_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, x_5 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, x_6 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$y_1 = 1, y_2 = -1, y_3 = 1, y_4 = -1, y_5 = 1, y_6 = -1$$

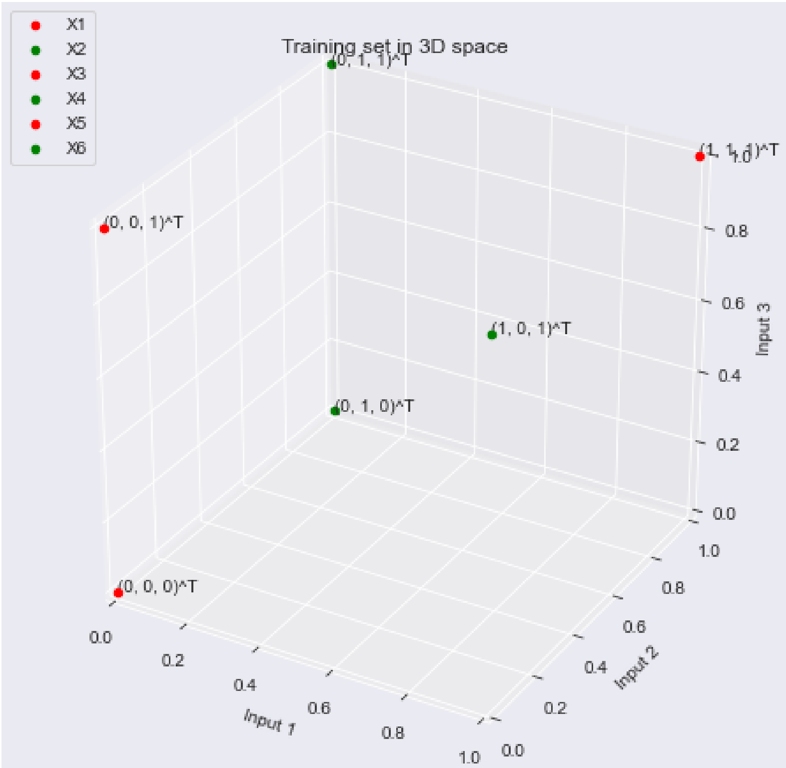


Figure 2: Training set in 3D space

03

$$y = (w_0 + w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9) \begin{pmatrix} z_0 = 1 \\ z_1 \\ z_2 \\ z_3 \\ z_1 z_2 \\ z_2 z_3 \\ z_1 z_3 \\ z_1^2 \\ z_2 \\ z_3^2 \end{pmatrix}$$

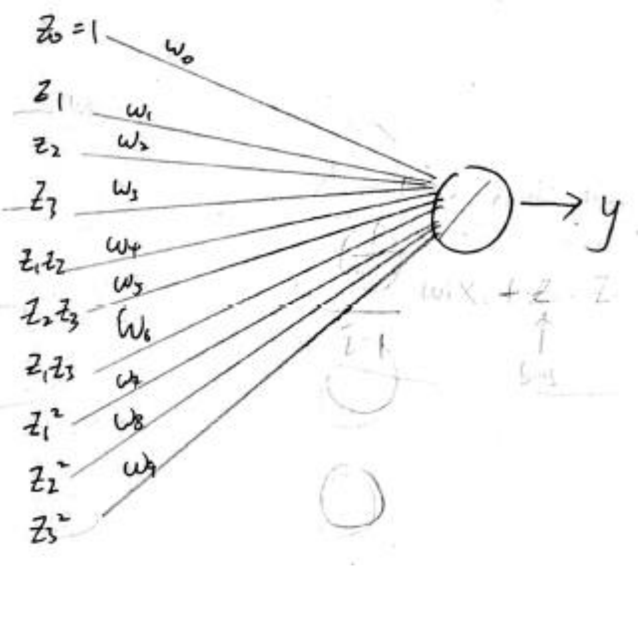
$$y = wz$$

$$y = w_0 + w_1 z_1 + w_2 z_2 + w_3 z_3 + w_4 z_1 z_2 + w_5 z_2 z_3 + w_6 z_1 z_3 + w_7 z_1^2 + w_8 z_2^2 + w_9 z_3^2$$

The dataset can be transformed as shown in table below.

x_k	Z_0	Z_1	Z_2	Z_3	Z_1Z_2	Z_2Z_3	Z_1Z_3	Z_1^2	Z_2^2	Z_3^3	y
X_1	1	0	0	0	0	0	0	0	0	0	1
X_2	1	0	1	1	0	1	0	0	1	1	-1
X_3	1	1	1	1	1	1	1	1	1	1	1
X_4	1	1	0	1	0	0	1	1	0	1	-1
X_5	1	0	0	1	0	0	0	0	0	1	1
X_6	1	0	1	0	0	0	0	0	1	0	-1

If transform the data from original space to feature space, then this data can be classified (



→ According to the gradient descent, we initialize w.

→ Then, we update the w as:

$$w_{t+1} = w_t + Lr \cdot (Y - \hat{Y}) x_i$$

→ After the initialization of weights update of weight, and result is optimized (shown in ipynb file)

The rest of answer of the question 3 can be found in the ipynb file.

4. The objective function of the perceptron can be defined differently as follows. Show the process of differentiating this expression, and using the differentiation result, present the weight update rule as the following delta rule.

Delta Rule: 델타 규칙: $\mathbf{w} = \mathbf{w} + \rho \sum_{\mathbf{x}_k \in Y} y_k \mathbf{x}_k$

Objective function $J(w) = \sum_{i=1}^{n} (y_i - \tau(w^T x_i)^2$

Q4 : Objective function $J(w) = \sum_{i=1}^{n} \{y_i - \tau(w^T x_i)\}^2$
(loss function)

$$\frac{\Delta J(w)}{\Delta w_i} = \frac{\Delta\{y_k - \tau(w_1 x_{k1} + w_2 x_{k2} + \ldots w_i x_{ki} + w_n x_{kn})\}^2}{\Delta w_i}$$

chain Rule.

$$\frac{\Delta J(w)}{\Delta w_i} = \frac{\Delta\{y_k - \tau(w_1 x_{k1} + \ldots w_{x_n} x_{kn})\}^2}{\Delta y_k - \tau(w_1 x_{k1} \ldots)} \times \frac{\Delta y_k - \tau(w_i x_{k1} \ldots)}{\Delta(w_1 x_{k1} \ldots)} \times \frac{\Delta(w_1 x_{k1} \ldots)}{\Delta w_1 x_{k1} \ldots}$$

$$\times \frac{\Delta w_1 x_{k1} + \ldots + w_i x_{ki} + \ldots}{\Delta w_i}$$

$$= 2\{y_k - \tau(w_1 x_{k1} + \ldots)\} \times (-1) \times \tau'(w_i x_{k1} + \ldots) \times x_{ki}$$

$$\therefore \frac{\Delta J(w)}{\Delta w_i} = \sum_{x_k \in Y} -2x_{ki} \left(y_k - \tau(w_1 x_{k1} + \ldots)\right) \tau'(w_1 x_{k1} + \ldots)$$

Delta rule
(델타 규칙)

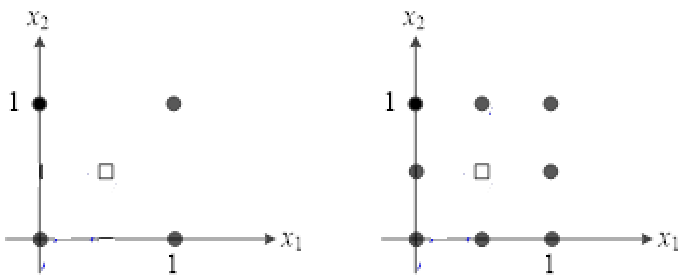
$$w_i = w_i - \rho \frac{\partial J(w)}{\partial w_i} \quad ; \rho: \text{learning rate}$$

$$w = w - \rho \sum_{x_k \in Y} -2x_{ki}(y_k - \tau(w_1 x_{k1} + \ldots))\tau'(w_1 x_{k1} + \ldots)$$

$$w = w + \rho \sum_{x_k \in Y} y_k x_k$$

※ proven.

5. Answer to the following classification problem



(1) Present each multi−layered perceptron that solves this classification problem.

(2) Answer whether the right situation can be solved with a perceptron with only two hidden nodes.

(1) The answer of (1) is in ipynb file of attached file.

(2) The right situation cannot be solved with a perceptron with a perceptron with only two hidden nodes, But, it can be solved by 3 hidden nodes. This can be explained as the Neural Network with 3 hidden layers can construct higher dimensions model that can represent the classifier to solve the difficult classification problem.

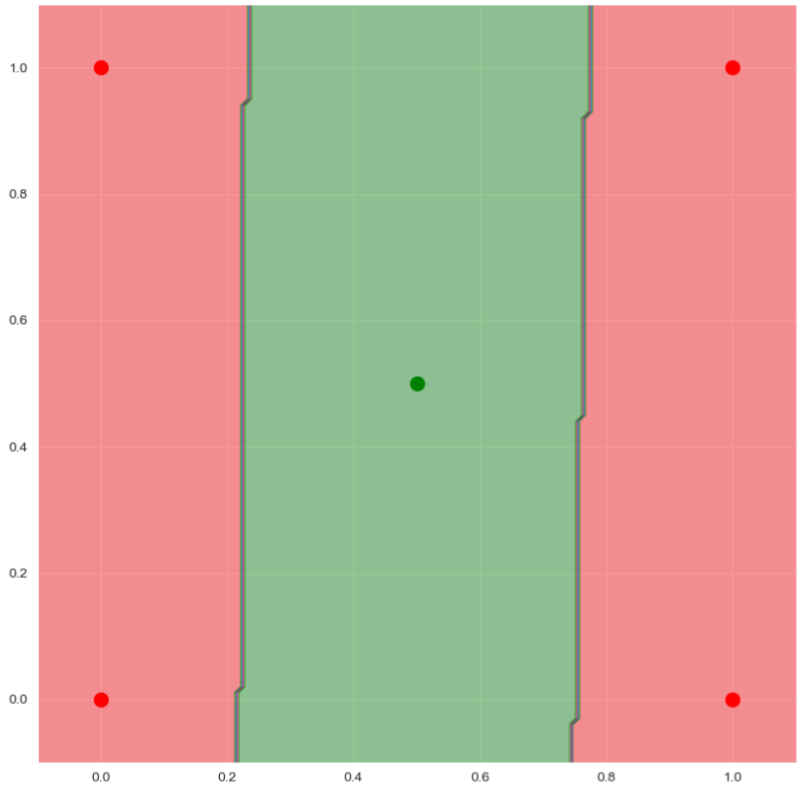


Figure 2: Classification of the left situation using 1 hidden layer with 2 hidden nodes

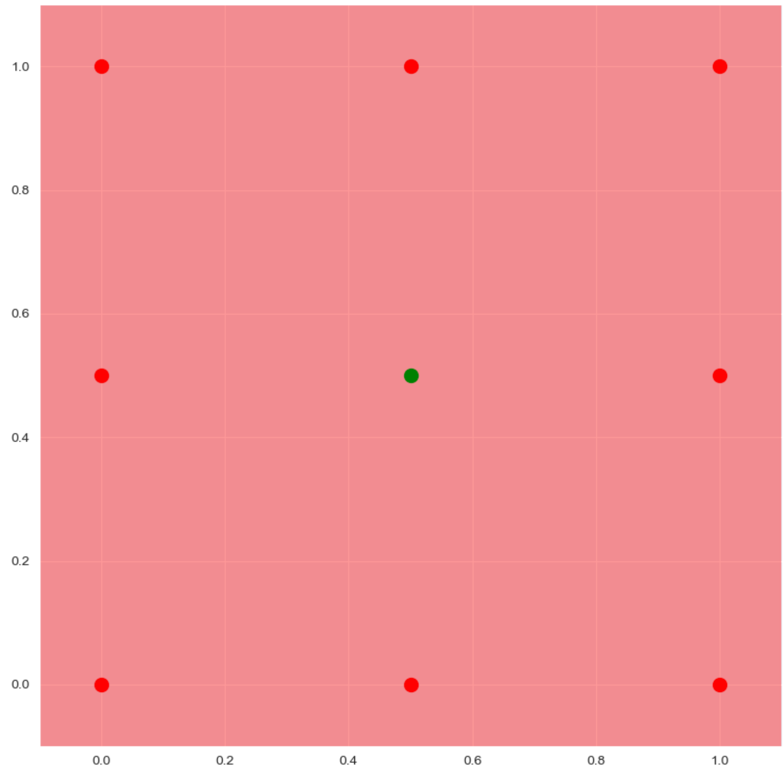


Figure 3: Classification of right situation using 1 hidden layer with 2 hidden nodes

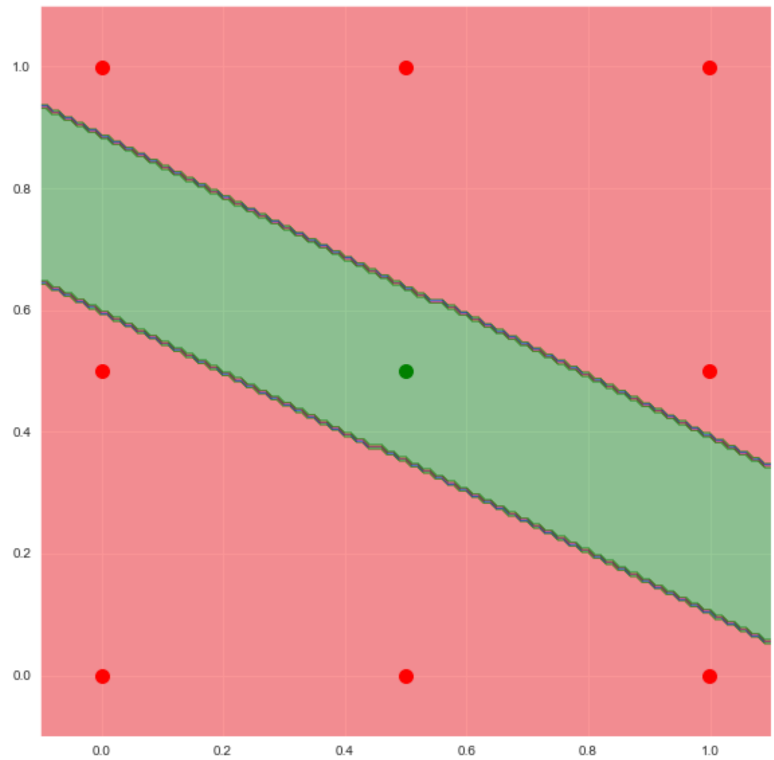


Figure 4: Classification of right situation using 1 hidden layer with 3 hidden nodes