Machine Learning
Assignment 3

Name: Aditya Aggarwal
Roll No.: 2012028

PROBLEM 1

assuming initial weights as

$w_1 = 1 \quad , \quad b_1 = 1$

$w_2 = 2 \quad , \quad b_2 = 0$

learning rate $\eta = 0.01$

Forward Pass: $\text{Relu}\,(\overbrace{w_1 x_1 + b_1}^{z_1})(w_2) + b_2 = z_2$

error $= \frac{1}{2}(y - y_p)^2 = E$

backpropagation:

$\Delta w_2 = -\eta \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta w_1}$ $\quad\quad$ $\Delta b_2 = -\eta \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta b_2}$

$\Rightarrow \Delta w_2 = \eta(y - y_p)\,\text{relu}(w_1 x + b_1)$ $\quad$ $\Delta b_2 = \eta(y - y_p)$

$\Delta w_1 = -\eta \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta \text{relu}(z_1)} \cdot \frac{\delta \text{relu}(z_2)}{\delta z_1} \cdot \frac{\delta z_1}{\delta w_1}$

$\Delta b_1 = -\eta \frac{\delta E}{\delta z_2} \cdot \frac{\delta z_2}{\delta \text{relu}(z_1)} \cdot \frac{\delta \text{relu}(z_1)}{\delta (z_1)} \cdot \frac{\delta z_1}{\delta b_1}$

Prediction $= (1(1) + 1) \cdot 2 + 0 = 4$

backward: $w_2 := w_2 + \Delta w_2$

$\Rightarrow w_2 = 2\cancel{4} + 0.01 \cdot (3 - 4)(2)$

$= 1.98$

$b_2 := b_2 + \Delta b_2$

$\Rightarrow b_2 = 0 + 0.01(-1)$

$= -0.01$

$w_1 := w_1 + \Delta w_1$

$w_1 = 1 + 0.01(3-4)(1.98)(1)(1)$

$= 0.9802$

$b_1 := b_1 + \Delta b_1$

$b_1 = 1 + (0.01)(3-4)(1.98)(1)(1)$

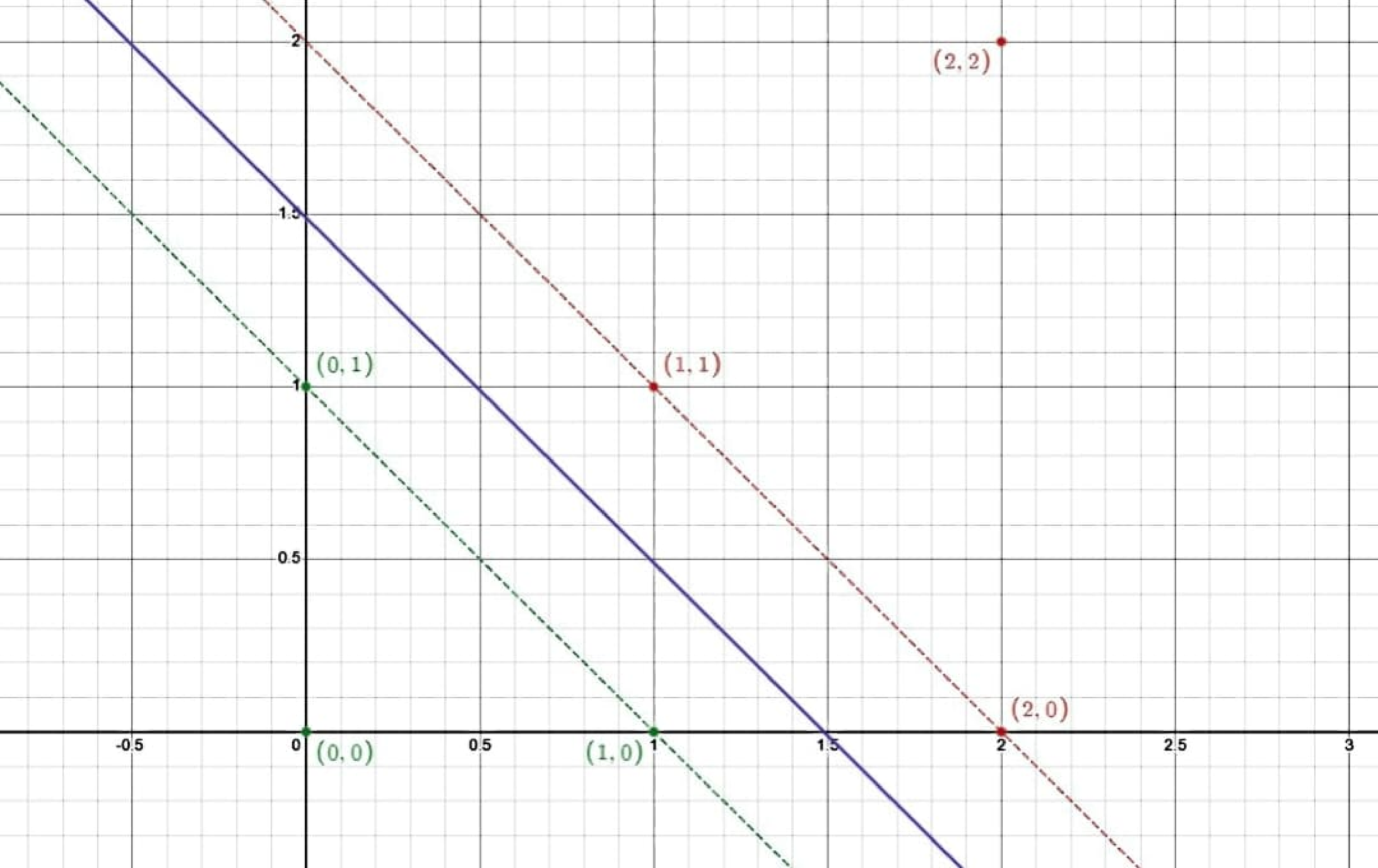
$= 0.9802$

After Sample 1 : $w_1 = 0.9802$ $b_1 = 0.9802$

$w_2 = 1.98$ $b_2 = -0.01$

Applying the same process for sample 2 & 3 :

After sample 2 and sample 3 } $\boxed{\begin{aligned} w_1 &= 0.8762, & b_1 &= 0.7914 \\ w_2 &= 1.9564, & b_2 &= -0.2475 \end{aligned}}$



2
1.5
1
0.5
-0.5
0
0.5
1
1.5
2
2.5
3
(2, 2)
(0, 1)
(1, 1)
(2, 0)
(0, 0)
(1, 0)

PROBLEM-2

(a) From the given datapoints, we plotted a graph using desmos graphing calculator.

From the plotted graph, it is clear that the data is linearly separable.

(b) condition for optimal margin : $y(w^T x + b) = 1$
from the graph and also from this condition, we conclude that $(1,0)$; $(0,1)$; $(1,1)$ are the support vectors.

$\Rightarrow \quad w_1 + b = 1$

and $\quad w_2 + b = 1$

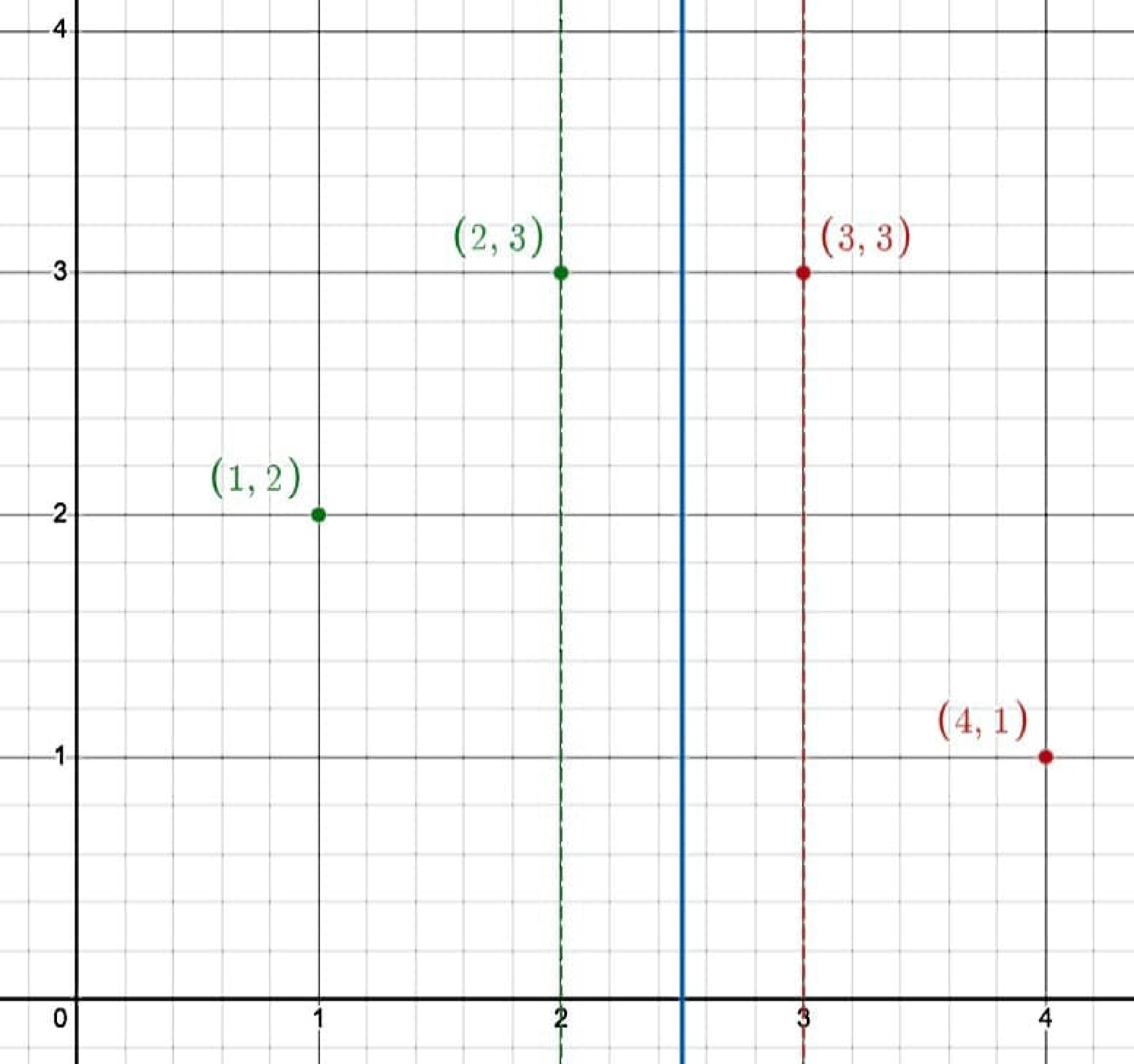
and $\quad w_1 + w_2 + b = -1$

Simultaneously solving these equations, we obtain w_1 as -2, w_2 as -2 & b as 3

$w^T x + b = 0 \quad \equiv -2x_1 - 2x_2 + 3 = 0 \equiv 2x_1 + 2x_2 - 3 = 0$
(making weights positive)

$\therefore \quad w_1 = w_2 = 2, \; b = -3$

maximum margin hyperplane $\equiv \boxed{2x_1 + 2x_2 - 3 = 0}$



(2, 3)
(3, 3)
(1, 2)
(4, 1)
4
3
2
1
0
1
2
3
4

PROBLEM-3

(b) given $W_1 = -2$, $W_2 = 0$ and $b = 5$, prediction for each point by the SVM $w_1 x_1 + w_2 x_2 + b = 0$:

$P_1 \equiv (-2)(1) + (0)(2) + 5 = 3$

$P_2 \equiv (-2)(2) + (0)(3) + 5 = 1$

$P_3 \equiv (-2)(3) + (0)(3) + 5 = -1$

$P_4 \equiv (-2)(4) + (0)(1) + 5 = -3$

It is clear that P_2 and P_3 are the support vectors as they are closest to the classifier.

$\therefore$ support vectors : P_2 & P_3

(a) The margin is calculated as $\frac{1}{\|w\|} = \frac{1}{\sqrt{w_1^2 + w_2^2}}$

$$= \frac{1}{\sqrt{(-2)^2 + 0^2}}$$

$$= 0.5$$

$\therefore$ The margin is 0.5

(c) Classification for (1, 3) :

$$y_{(1,3)} = \text{sgn}(w^T x + b)$$

$$w^T x + b = \begin{bmatrix} -2 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + 5$$

$$y_{(1,3)} = \text{sgn}(3) = +1$$

$\therefore$ (1, 3) is assigned the label +1