<AML Homework 4>

Matheus Clemente Bafutto (mc2722)
Sungseo Park (sp2528)

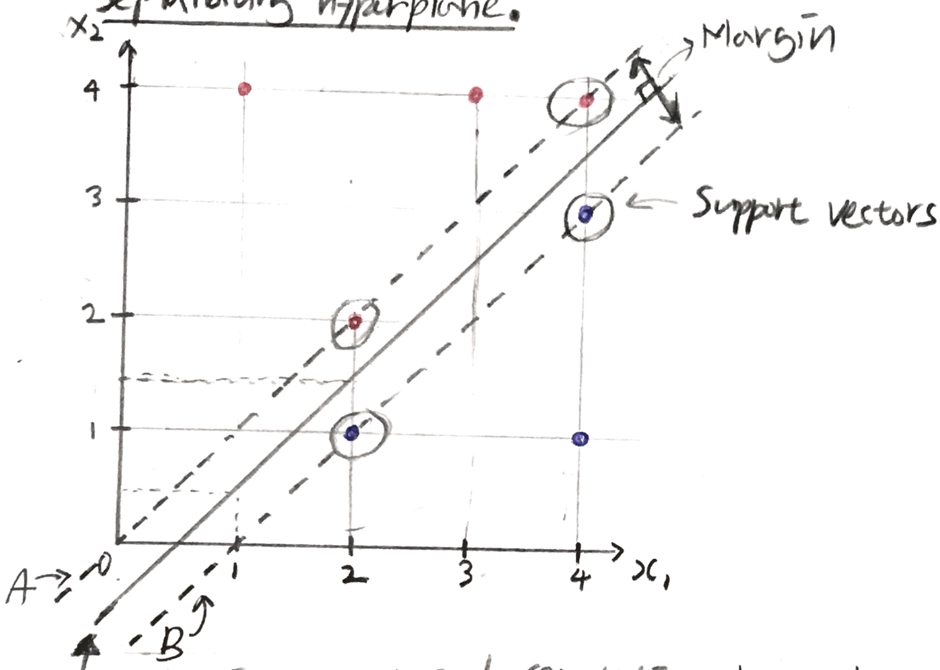
- Written Exercises

Q1. Maximum-margin classifier

Given $n=7$, observation in $p=2$ dimension

X_1	X_2	Y
3	4	Red
2	2	Red
4	4	Red
1	4	Red
2	1	Blue
4	3	Blue
4	1	Blue

a) Sketch the observation and maximum-margin separating hyperplane.



b) Describe the classification rule for the maximal margin classifier.

- Since the line goes through $(2, 1.5)$ and $(4, 3.5)$

$x_2 = m x_1 + c$, $m = \dfrac{3.5-1.5}{4-2} = 1 \Rightarrow x_2 = x_1 + C$.

If $x_1 = 2$ and $x_2 = 1.5$, $1.5 = 2 + C$ and $C = -0.5$. Then $0.5 - x_1 + x_2 = 0$

Therefore, the classification rule:

"classify to Red if $0.5 - x_1 + x_2 > 0$, and classify to Blue otherwise."

Thus, $\beta_0 = 0.5$, $\beta_1 = -1$ and $\beta_2 = 1$.

(c) On your sketch, indicate the margin for the maximal margin hyperplane.

The margin for the maximal margin hyperplane is the perpendicular distance from points on the two dashed lines.

- Line A goes through (2,2) and (4,4). $x_2 = mx_1 + c$

  $m = \frac{4-2}{4-2} = 1 \Rightarrow x_2 = x_1 + c$. When $x_1 = 2$ and $x_2 = 2$, then $c = 0$

  Thus, Line A $\Rightarrow x_2 = x_1$

- Line B goes through (2,1) and (4,3).

  $m = \frac{3-1}{4-2} = 1 \Rightarrow x_2 = x_1 + c$. when $x_1 = 2$ and $x_2 = 1$, then $c = -1$

  Thus, Line B $\Rightarrow x_2 = x_1 - 1$

- The formula for the distance between the two parallel lines:

  $d = \frac{|c_1 - c_2|}{\sqrt{1+m^2}}$. Then $d = \frac{|0+1|}{\sqrt{1+1}} = \frac{1}{\sqrt{2}}$

  Therefore, the length of the margin is $\frac{1}{\sqrt{2}}$

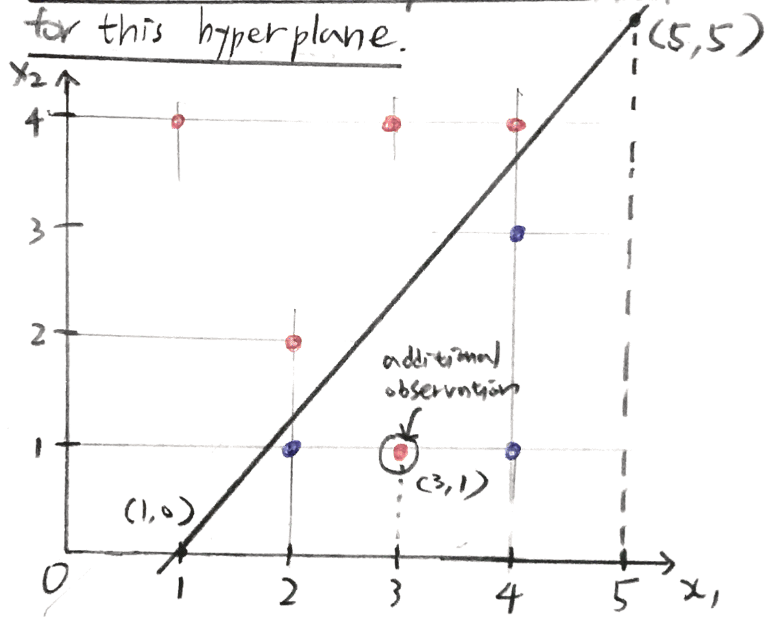
(d) Indicate the support vectors for the maximal margin classifier.

- The support vectors are (2,1), (2,2), (4,3) and (4,4).

(e) Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.

- The maximal margin hyperplane would be only affected by the support vectors. Since the seventh observation is not a support vector, it would not affect the maximal margin hyperplane.

(f) Sketch a hyperplane that separates the data, but is not the maximum-margin separating hyperplane. Provide the equation for this hyperplane.



- The line goes through (1,0) and (5,5). Then the slope would be $m = \frac{5-0}{5-1} = 1.25$. Given $x_2 = mx_1 + C$ and the point (1,0), $0 = (1.25)(1) + C$. Thus $C = -1.25$. Therefore, the equation for this hyperplane would be

$$y = 1.25x - 1.25$$

(g) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

- When the red point (3,1) is added to the plot, the two classes are no longer separable by a hyperplane.

Written Exercises

$\frac{1}{3} \cdot 2 + b = 2$

$b = 2 - \frac{2}{3} = \frac{4}{3}$

$6 \cdot \frac{-5}{3} + b = 5$

$b = 5 + 10 = 15$

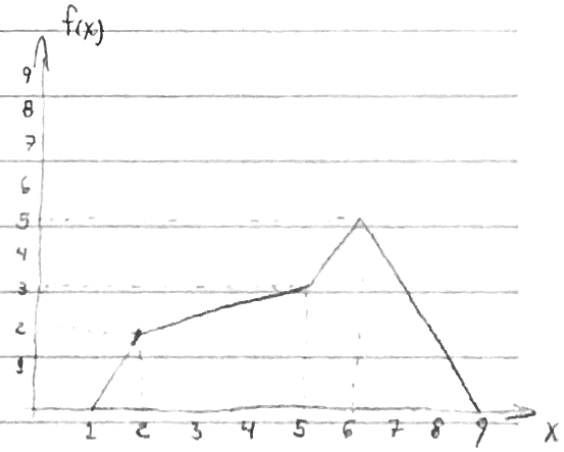
2.

Pride Rock function:

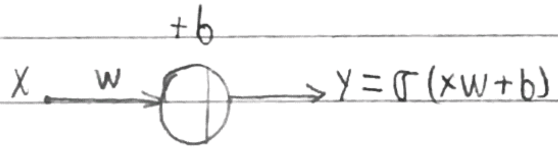
$$f(x) = \begin{cases} 0, & x < 1 \\ 2x-2, & 1 < x < 2 \\ \frac{1}{3}x + \frac{4}{3}, & 2 < x < 5 \\ 2x - 7, & 5 < x < 6 \\ \frac{-5}{3}x + 15, & 6 < x < 9 \\ 0, & x > 9 \end{cases}$$

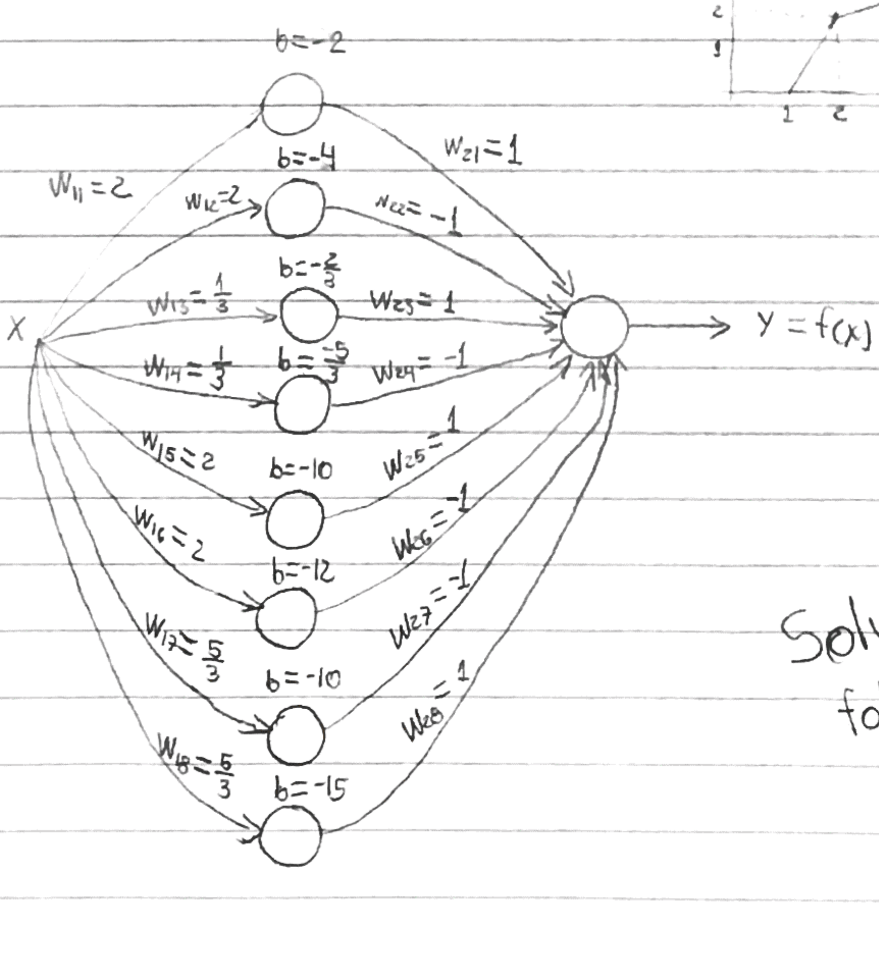
ReLU function:

$$\sigma(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



1 unit with reLU

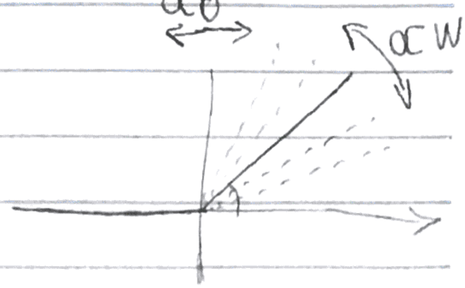




Solution on following pages

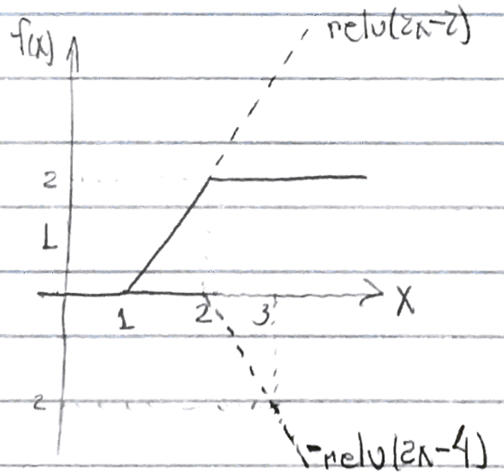
The pride rock function is a sum of ramps. we can obtain this function by summing relu's with different biases and weights.

$$\sigma(wx+b) = \begin{cases} xw + bw, & xw + bw > 0 \\ 0, & \text{otherwise} \end{cases}$$



We can obtain each ramp behavior with an instance of a relu $\sigma(wx+b)$. To tackle overlaps between the relus, we can timely introduce a twin relu for each of the ramps in order to cancel them off.

ex: $f(x) = relu(2x-2) - relu(2x-4)$

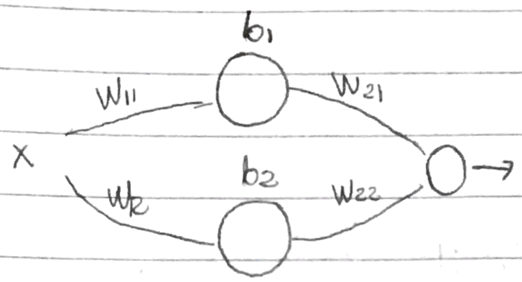


In this specific example, the relus increase at same rate after introduction of mon -relu(2x-4).

Still on this regard, we can repeat this procedure for the remaining ramps of pride rock. By timing the biases and matching weights, it is possible to recreate pride rock function using 4 pairs of relus such as those bellow

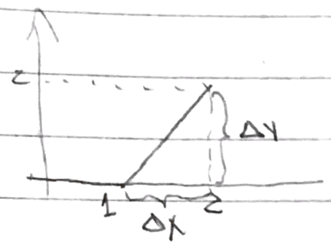
$$\text{Pride Rock}(x) = relu(2x-2) - relu(2x-4) + relu\left(\frac{x}{3} - \frac{2}{3}\right) - relu\left(\frac{x}{3} - \frac{5}{3}\right) + relu(2x-10) - relu(2x-12) + relu\left(\frac{5}{3}x - 10\right) - relu\left(\frac{5}{3}x - 15\right)$$

In sumary, to get a ramp in an interval, we take 2 units (1 for ramp and another for stopping), align weights and time their biases like this:



$w_{11} = w_{12} = \frac{\Delta y}{\Delta x} = 2$ ($\Delta y \rightarrow y_1 - y_0$, $\Delta x \rightarrow x_1 - x_0$)

$b_1 = -w_{11} \cdot x_0$
$b_2 = -w_{12} \cdot x_1$



$w_{21} = 1$
$w_{22} = -1$

For Pride Rock, calculation summary:

	w_{11}	w_{12}	b_1	b_2	w_{21}	w_{22}
$1 \leq x < 2$	$\frac{2}{1} = 2$	$\frac{2}{1} = 2$	$-2 \cdot 1 = -2$	$-2 \cdot 2 = -4$	1	-1
$2 \leq x < 5$	$\frac{3-2}{5-2} = 1/3$	$\frac{1}{3}$	$-2 \cdot \frac{1}{3} = -2/3$	$-5 \cdot \frac{1}{3} = -5/3$	1	-1
$5 \leq x < 6$	$\frac{5-3}{6-5} = 2$	2	$-5 \cdot 2 = -10$	$-6 \cdot 2 = -12$	1	-1
$6 \leq x < 9$	$\frac{5-0}{9-6}$* $= 5/3$	$\frac{5}{3}$	$-\frac{5}{3} \cdot 6 = -10$	$-\frac{5}{3} \cdot 9 = -15$	-1	1 *

* Switched the minus sign here to simplify calculations.

Reference:
- Neuralnetworksanddeeplearning.com/chap4.html