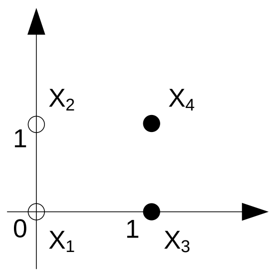
CSE 5526: Homework 2 Solution

1.
a) (2 points)



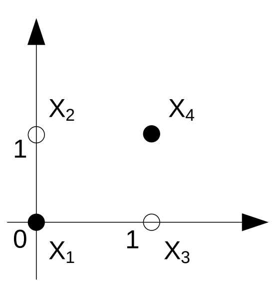
We set the desired output of points in C_1 to -1, and that of points in C_2 to 1.
Training in action (Assume in augmented space with η=1/2)
$w(n+1)=w(n)+\Delta w(n)=w(n)+\eta[d(n)-y(n)]x(n)$

Weight (**w**)	Input (**x**)	Desired (d)	Actual output (y)	($d-y$)
(0, 0, 0)	(0, 0, 1)	-1	1	-2
(0, 0, -1)	(0, 1, 1)	-1	-1	0
(0, 0, -1)	(1, 0, 1)	1	-1	2
(1, 0, 0)	(1, 1, 1)	1	1	0
(1, 0, 0)	(0, 0, 1)	-1	1	-2
(1, 0, -1)	(0, 1, 1)	-1	-1	0
(1, 0, -1)	(1, 0, 1)	1	1	0
(1, 0, -1)	(1, 1, 1)	1	1	0
(1, 0, -1)	(0, 0, 1)	-1	-1	0

correct classification

Hence, the resulting decision boundary is $\mathbf{w}^T\mathbf{x}=0$, therefore, $x_1-1=0$ or $x_1=1$.

b) (1 point)
Here we set the desired output of points in C_1 to 1, and that of points in C_2 to -1.

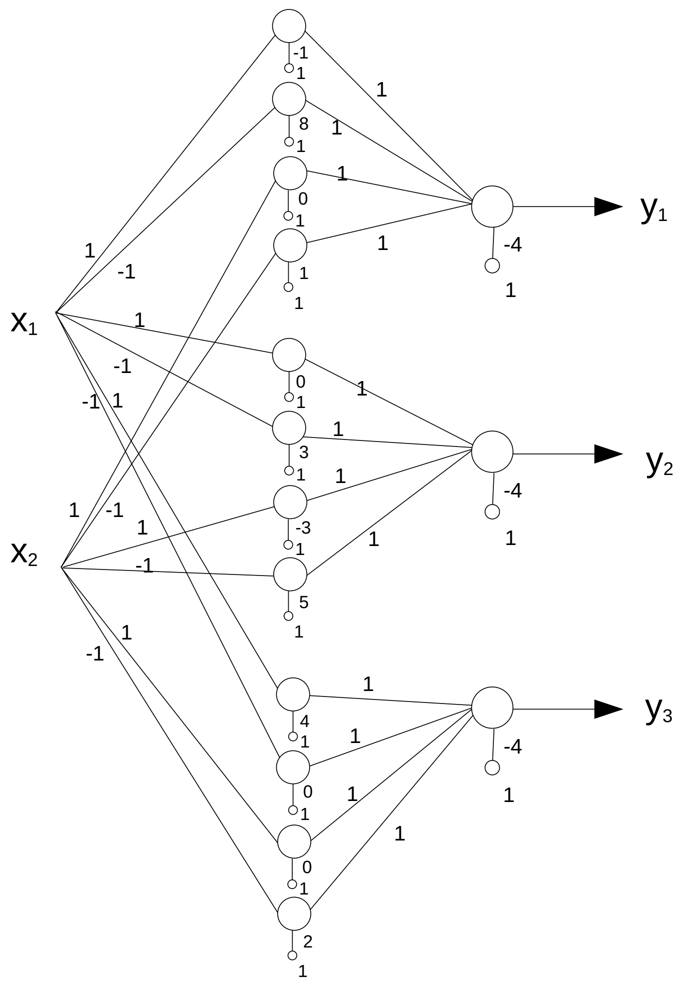


Training in action (Assume in augmented space with η=1/2)

Weight (**w**)	Input (**x**)	Desired (d)	Actual output (y)	($d-y$)
(0, 0, 0)	(0, 0, 1)	-1	1	-2
(0, 0, -1)	(0, 1, 1)	1	-1	2
(0, 1, 0)	(1, 0, 1)	1	1	0
(0, 1, 0)	(1, 1, 1)	-1	1	-2
(-1, 0, -1)	(0, 0, 1)	-1	-1	0
(-1, 0, -1)	(0, 1, 1)	1	-1	2
(-1, 1, 0)	(1, 0, 1)	1	-1	2
(0, 1, 1)	(1, 1, 1)	-1	1	-2
(-1, 0, 0)	(0, 0, 1)	-1	1	-2
(-1, 0, -1)	(0, 1, 1)	1	-1	2
(-1, 1, 0)	(1, 0, 1)	1	-1	2
(0, 1, 1)	(1, 1, 1)	-1	1	-2
(-1, 0, 0)	(0, 0, 1)	-1	1	-2
…	…	…	…	…

One can see from the above table that the weights will repeat. Perceptron learning thus will not converge but loop forever.

2. (3 points)



Explanation:
Consider the first group of 4 neurons in the first layer.
The first neuron will output 1 if x_1 >= 1, and -1 otherwise.
The second neuron will output 1 if - x_1 >= -8 (equivalently x_1 <= 8), and -1 otherwise.
The third neuron will output 1 if x_2 >= 0, and -1 otherwise.
The fourth neuron will output 1 if - x_2 >= -1 (equivalently x_2 <= 1), and -1 otherwise.

The first neuron in the second layer will output 1 if all four of the above neurons output 1, and will output -1 otherwise.

This gives the conditions 1 <= x_1 <= 8, and 0 <= x_2 <= 1, which is exactly the area occupied by Class 1.

The other 2 classes are determined similarly.

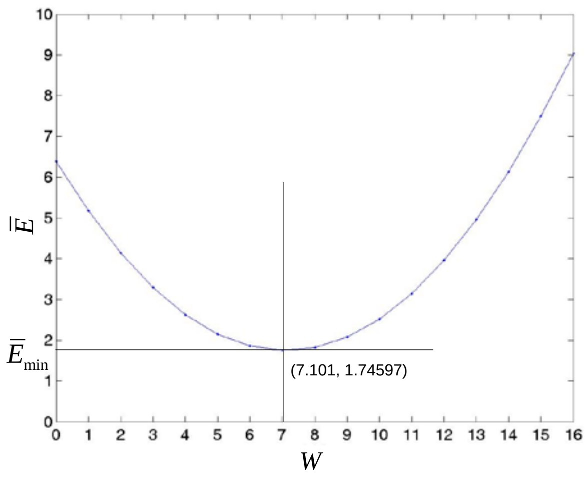
3. (4 points)

X = {-0.5, -0.2, -0.1, 0.3, 0.4, 0.5, 0.7}
D = {-1, 1, 2, 3.2, 3.5, 5, 6}

$$
\begin{aligned}
E &= \frac{1}{2}\sum_i (d_i - wx_i)^2 = \frac{1}{2}\sum_i (x_i{}^2 w^2 - 2d_i x_i w + d_i{}^2) \\
&= \frac{1}{2}[(-0.5)^2 w^2 - 2\cdot(-1)\cdot(-0.5)w + (-1)^2 \\
&\qquad +(-0.2)^2 w^2 - 2\cdot 1\cdot(-0.2)w + 1^2 \\
&\qquad +(-0.1)^2 w^2 - 2\cdot 2\cdot(-0.1)w + 2^2 \\
&\qquad +0.3^2 w^2 - 2\cdot 3.2\cdot 0.3w + 3.2^2 \\
&\qquad +0.4^2 w^2 - 2\cdot 3.5\cdot 0.4w + 3.5^2 \\
&\qquad +0.5^2 w^2 - 2\cdot 5\cdot 0.5w + 5^2 \\
&\qquad +0.7^2 w^2 - 2\cdot 6\cdot 0.7w + 6^2] \\
&= \frac{1}{2}[1.29w^2 - 18.32w + 89.49]
\end{aligned}
$$

Average over data points:

$$
\overline{E} = \frac{E}{7} = \frac{1.29w^2 - 18.32w + 89.49}{14}
$$



Direct differentiation:

$$
\frac{\partial \overline{E}}{\partial w} = \frac{2.58w - 18.32}{14}
$$

the gradient at the point w=2:

$$
\nabla E_w = \frac{\partial \overline{E}}{\partial w} = \frac{2.58\times 2 - 18.32}{14} = -0.94
$$

LMS approximation:

$$\nabla E_w = \frac{\partial}{\partial w(k)} E = \frac{1}{2}\frac{\partial}{\partial w(k)}(e^2(k)) = -e(k)x(k)$$

when w=2

$x(k)$	$y(k)$	$d(k)$	$e(k)=d(k)-y(k)$	$-e(k)x(k)$
-0.5	-1	-1	0	0
-0.2	-0.4	1	1.4	0.28
-0.1	-0.2	2	2.2	0.22
0.3	0.6	3.2	2.6	-0.78
0.4	0.8	3.5	2.7	-1.08
0.5	1.0	5	4	-2
0.7	1.4	6	4.6	-3.22

So

$$\nabla \overline{E}_w(k) = \frac{1}{7}\left[0.0 + 0.28 + 0.22 - 0.78 - 1.08 - 2 - 3.22\right] = -0.94$$

Therefore, the gradients at the point w=2 from the 2 approaches agree exactly.