Practice-2

Machine Learning Techniques

Karthik Thiagarajan

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

Q-11

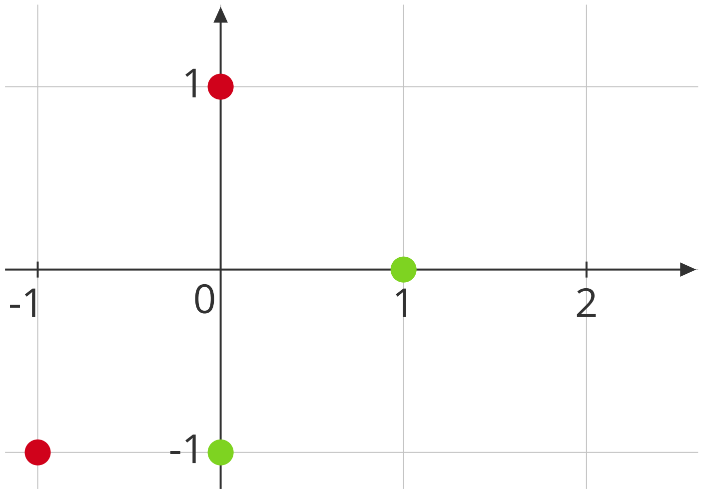
- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$

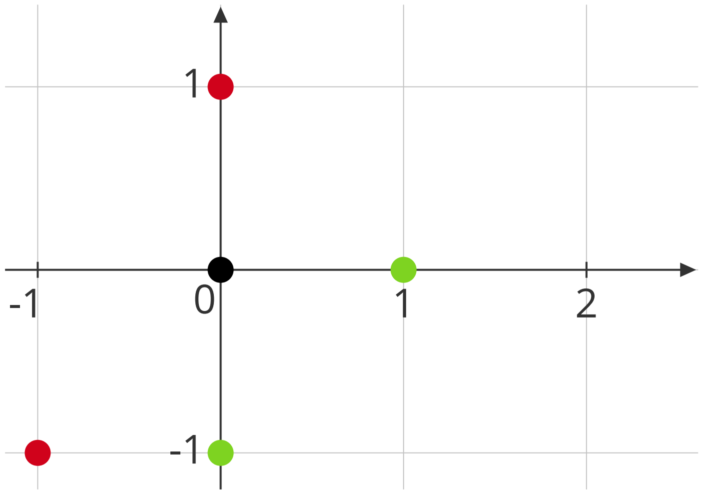


$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



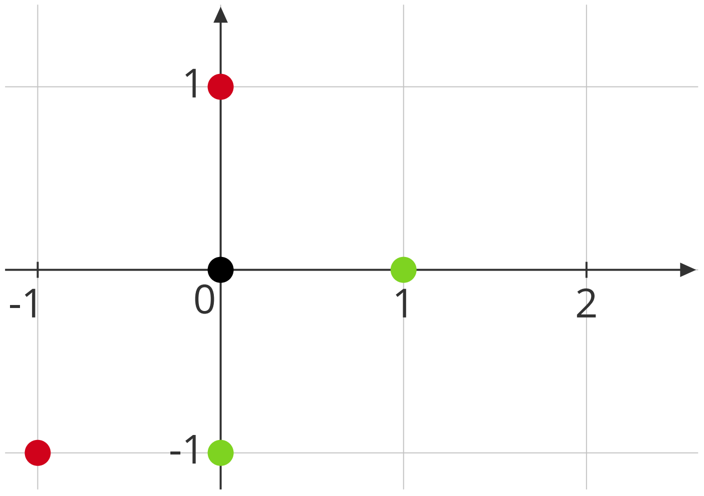
$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	
$(0, -1)$	1	
$(0, 1)$	-1	
$(-1, -1)$	-1	

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



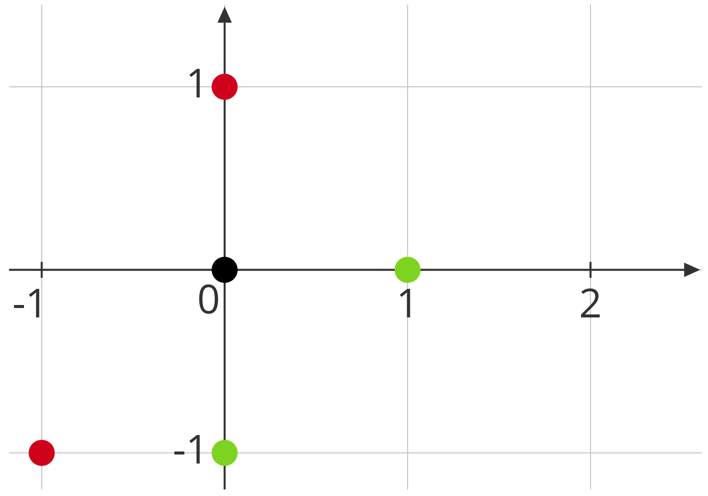
$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

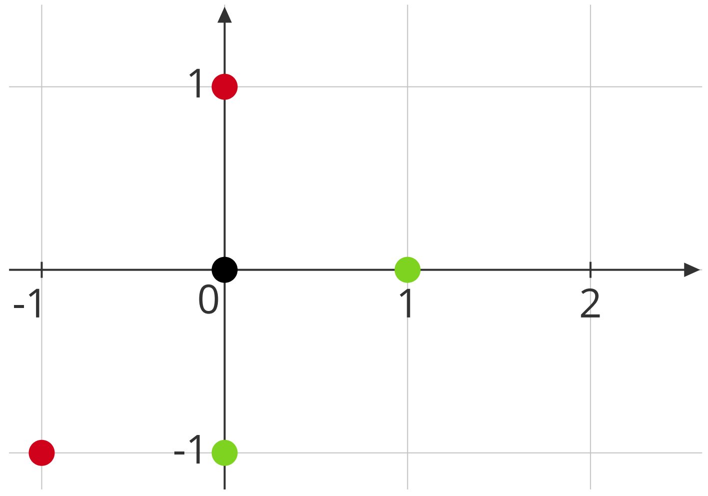
$$\mathbf{w}^1 = \mathbf{w}^0 + \mathbf{x}_i y_i$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	1

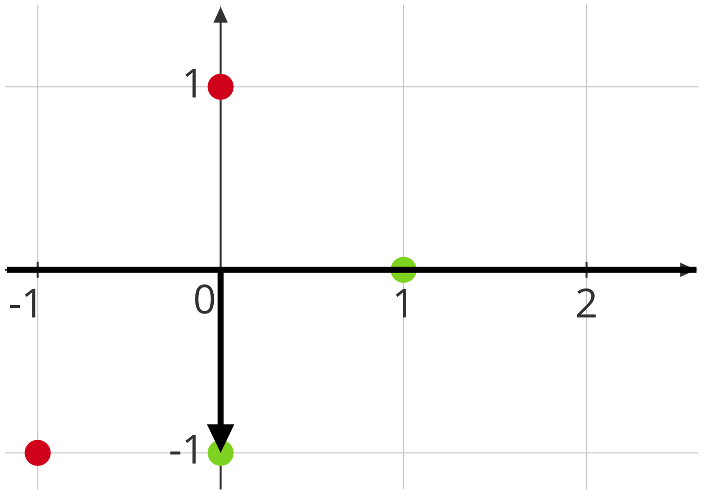
$$\begin{aligned} \mathbf{w}^1 &= \mathbf{w}^0 + \mathbf{x}_3 y_3 \\ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(-1) \\ &= \begin{bmatrix} 0 \\ -1 \end{bmatrix} \end{aligned}$$

$$\mathbf{w}^1 = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	1

$$\mathbf{w}^1 = \mathbf{w}^0 + \mathbf{x}_3 y_3$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (-1)$$

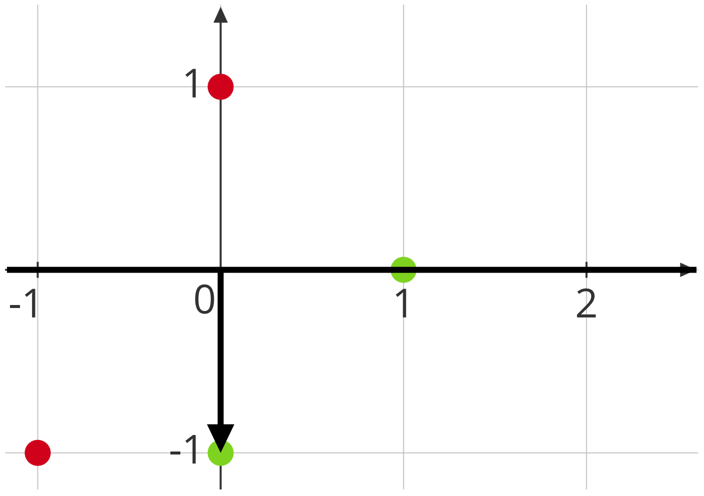
$$= \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

$$\mathbf{w}^1 = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



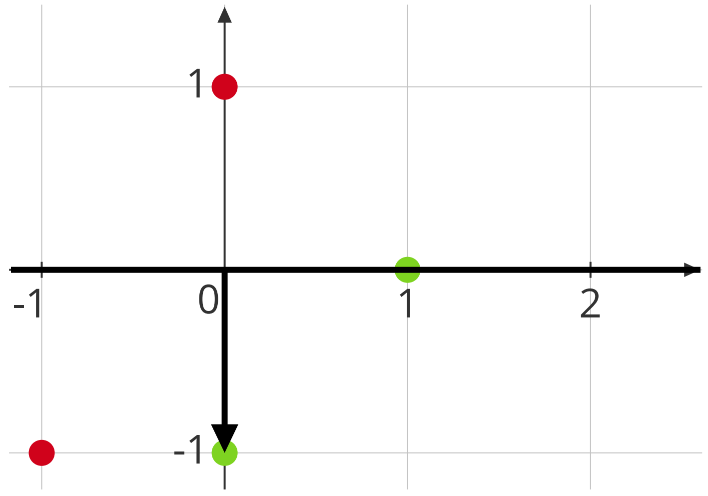
$$\mathbf{w}^1 = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	
$(0, -1)$	1	
$(0, 1)$	-1	
$(-1, -1)$	-1	

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



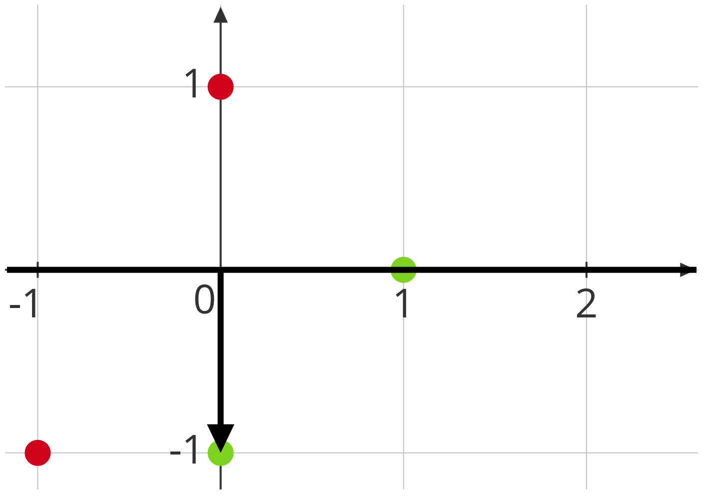
$$\mathbf{w}^1 = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^1 = \begin{bmatrix} 0 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	1

$$\mathbf{w}^2 = \mathbf{w}^1 + \mathbf{x}_4 y_4$$

$$= \begin{bmatrix} 0 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} (-1)$$

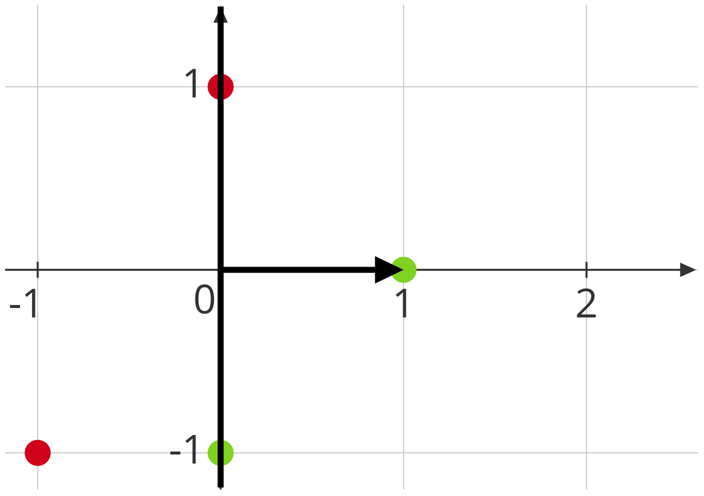
$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mathbf{w}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



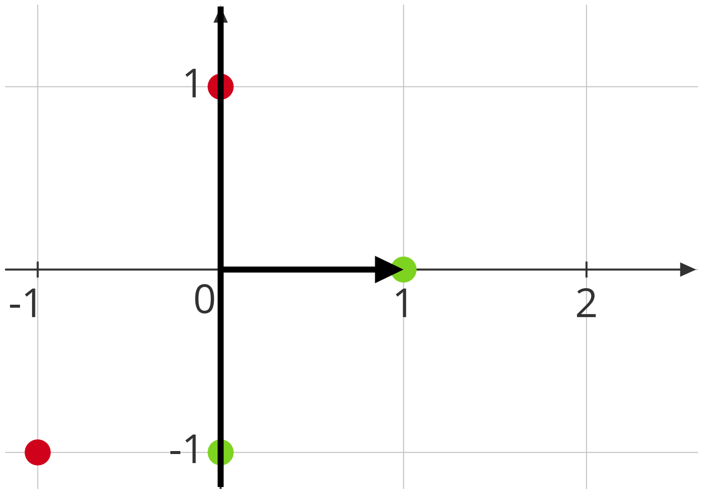
$$\mathbf{w}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	
$(0, -1)$	1	
$(0, 1)$	-1	
$(-1, -1)$	-1	

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



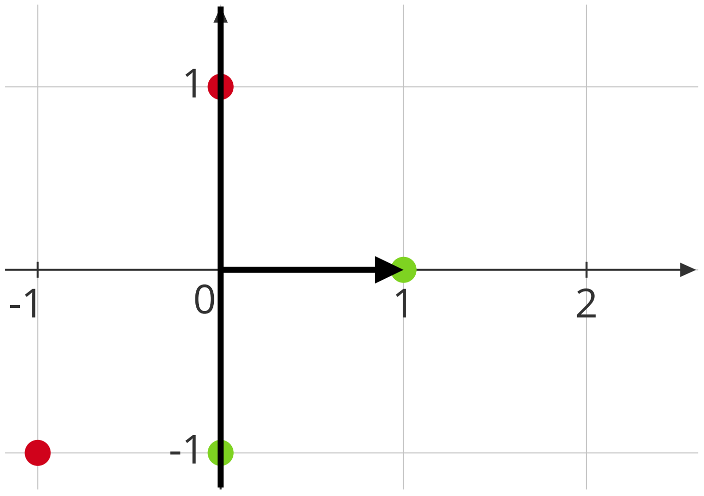
$$\mathbf{w}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	-1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	-1

$$\mathbf{w}^3 = \mathbf{w}^2 + \mathbf{x}_3 y_3$$

$$= \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(-1)$$

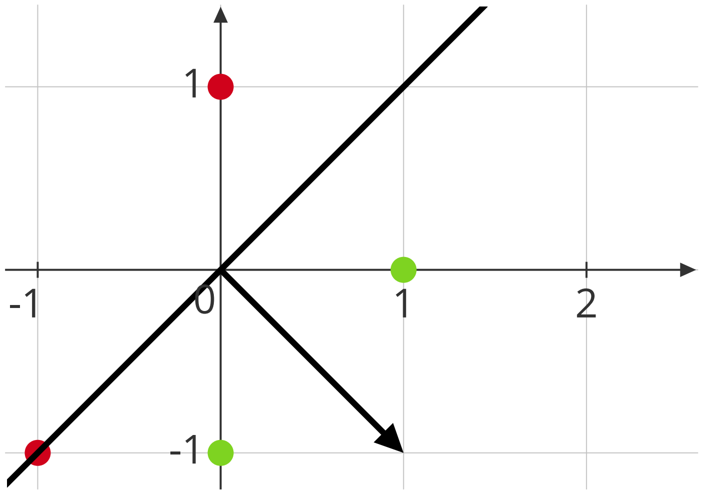
$$= \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\mathbf{w}^3 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



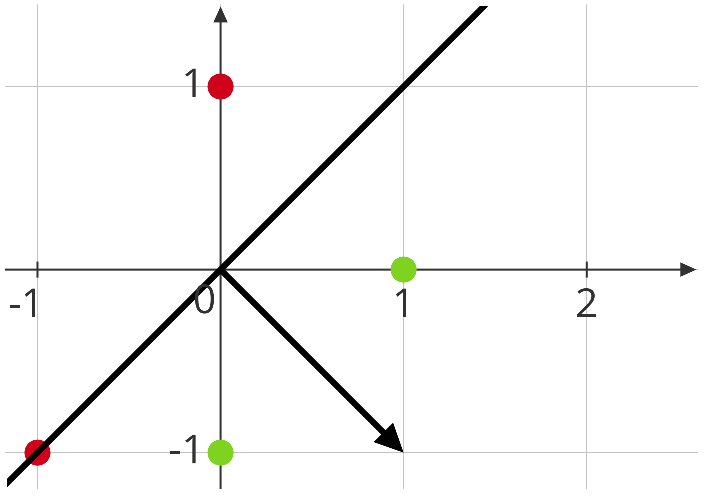
$$\mathbf{w}^3 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	
$(0, -1)$	1	
$(0, 1)$	-1	
$(-1, -1)$	-1	

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



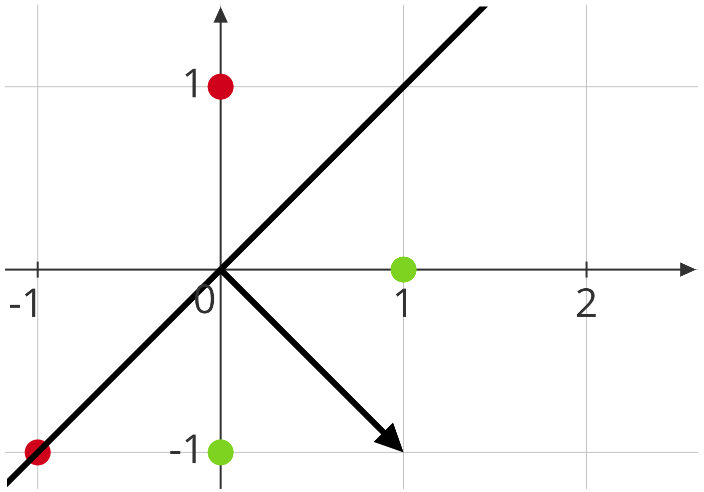
$$\mathbf{w}^3 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^3 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\hat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	1

$$\mathbf{w}^4 = \mathbf{w}^3 + \mathbf{x}_4 y_4$$

$$= \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix}(-1)$$

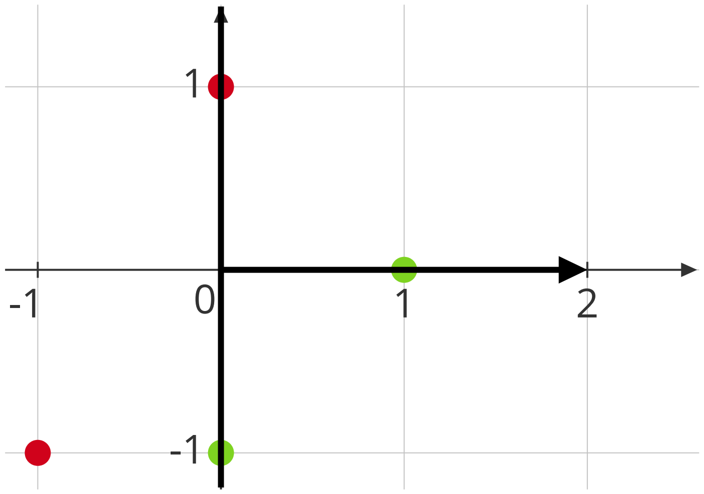
$$= \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\mathbf{w}^4 = \begin{bmatrix} 2 & 0 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = [\alpha_1 \ \cdots \ \alpha_n]^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = [1 \ 1 \ -1 \ -1]^T$$



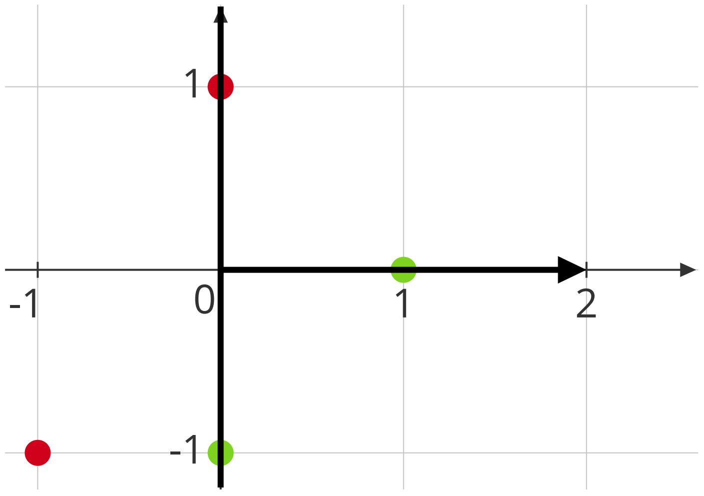
$$\mathbf{w}^4 = [2 \ 0]^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	
$(0, -1)$	1	
$(0, 1)$	-1	
$(-1, -1)$	-1	

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



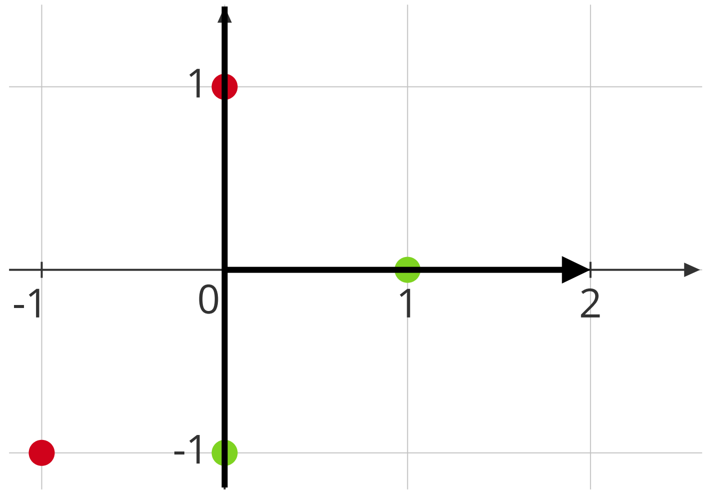
$$\mathbf{w}^4 = \begin{bmatrix} 2 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	-1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^4 = \begin{bmatrix} 2 & 0 \end{bmatrix}^T$$

$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	1
$(-1, -1)$	-1	-1

$$\mathbf{w}^5 = \mathbf{w}^4 + \mathbf{x}_3 y_3$$

$$= \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(-1)$$

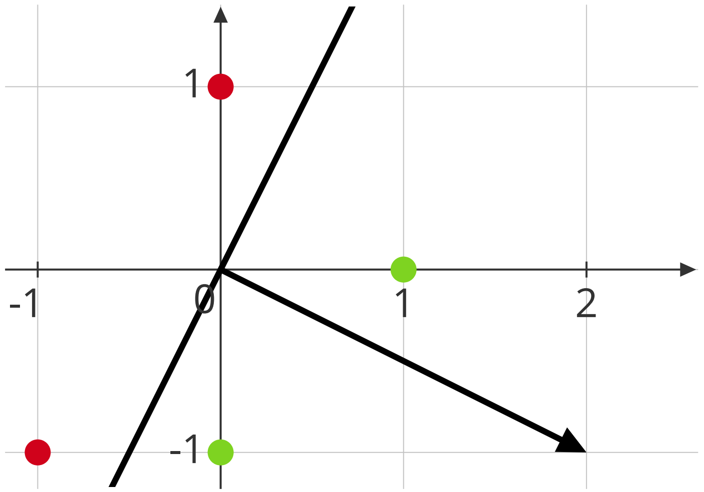
$$= \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

$$\mathbf{w}^5 = \begin{bmatrix} 2 & -1 \end{bmatrix}^T$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$



$$\mathbf{w}^5 = \begin{bmatrix} 2 & -1 \end{bmatrix}^T$$

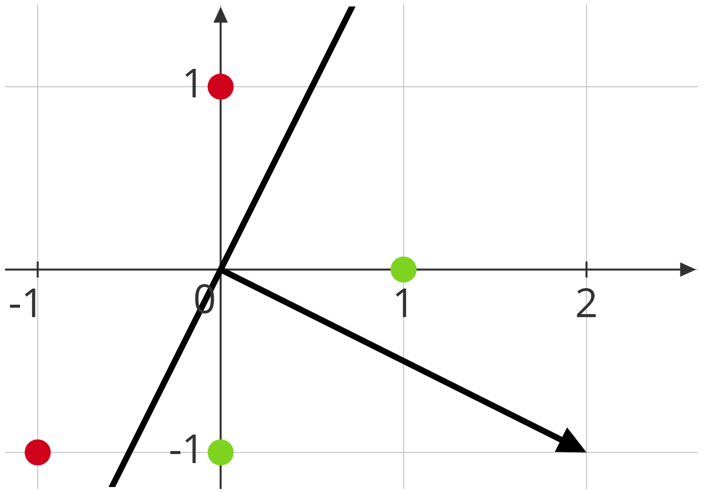
$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	-1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$

$$\mathbf{w}^* = \begin{bmatrix} 2 & -1 \end{bmatrix}^T$$



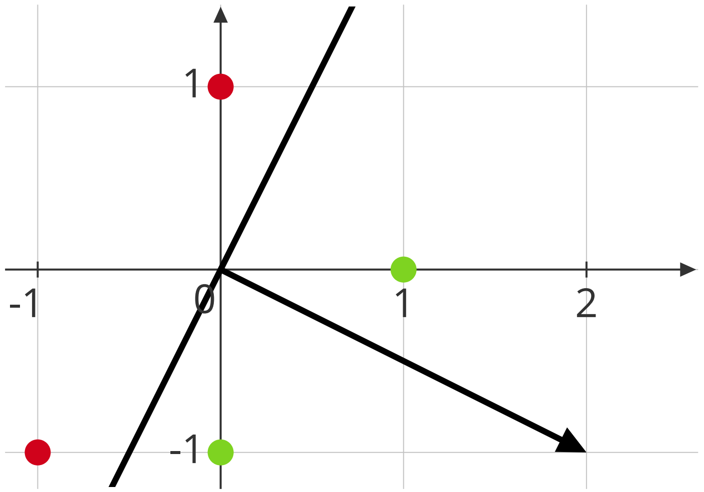
$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	-1

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$

$$\mathbf{w}^* = \begin{bmatrix} 2 & -1 \end{bmatrix}^T$$



$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	-1

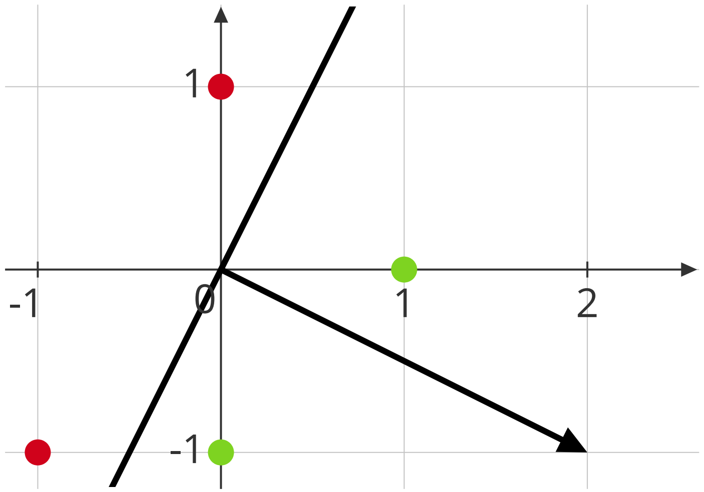
$$\begin{aligned} \mathbf{w}^* &= 0 \cdot x_1 y_1 + 0 \cdot x_2 y_2 + 3 \cdot \mathbf{x}_3 y_3 + 2 \cdot \mathbf{x}_4 y_4 \\ &= 3 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot (-1) + 2 \cdot \begin{bmatrix} -1 \\ -1 \end{bmatrix} \cdot (-1) \\ &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} \end{aligned}$$

Q-11

- Train a perceptron
- Cycle through the points from left to right
- If $\mathbf{w}^* = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i y_i$, find $\alpha = \begin{bmatrix} \alpha_1 & \cdots & \alpha_n \end{bmatrix}^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & -1 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$$

$$\mathbf{w}^* = \begin{bmatrix} 2 & -1 \end{bmatrix}^T$$



$\mathbf{x}$	y	$\widehat{y}$
$(1, 0)$	1	1
$(0, -1)$	1	1
$(0, 1)$	-1	-1
$(-1, -1)$	-1	-1

$$\begin{aligned} \mathbf{w}^* &= 0 \cdot x_1 y_1 + 0 \cdot x_2 y_2 + 3 \cdot \mathbf{x}_3 y_3 + 2 \cdot \mathbf{x}_4 y_4 \\ &= 3 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot (-1) + 2 \cdot \begin{bmatrix} -1 \\ -1 \end{bmatrix} \cdot (-1) \\ &= \begin{bmatrix} 2 \\ -1 \end{bmatrix} \end{aligned}$$

$$\alpha = \begin{bmatrix} 0 & 0 & 3 & 2 \end{bmatrix}^T$$

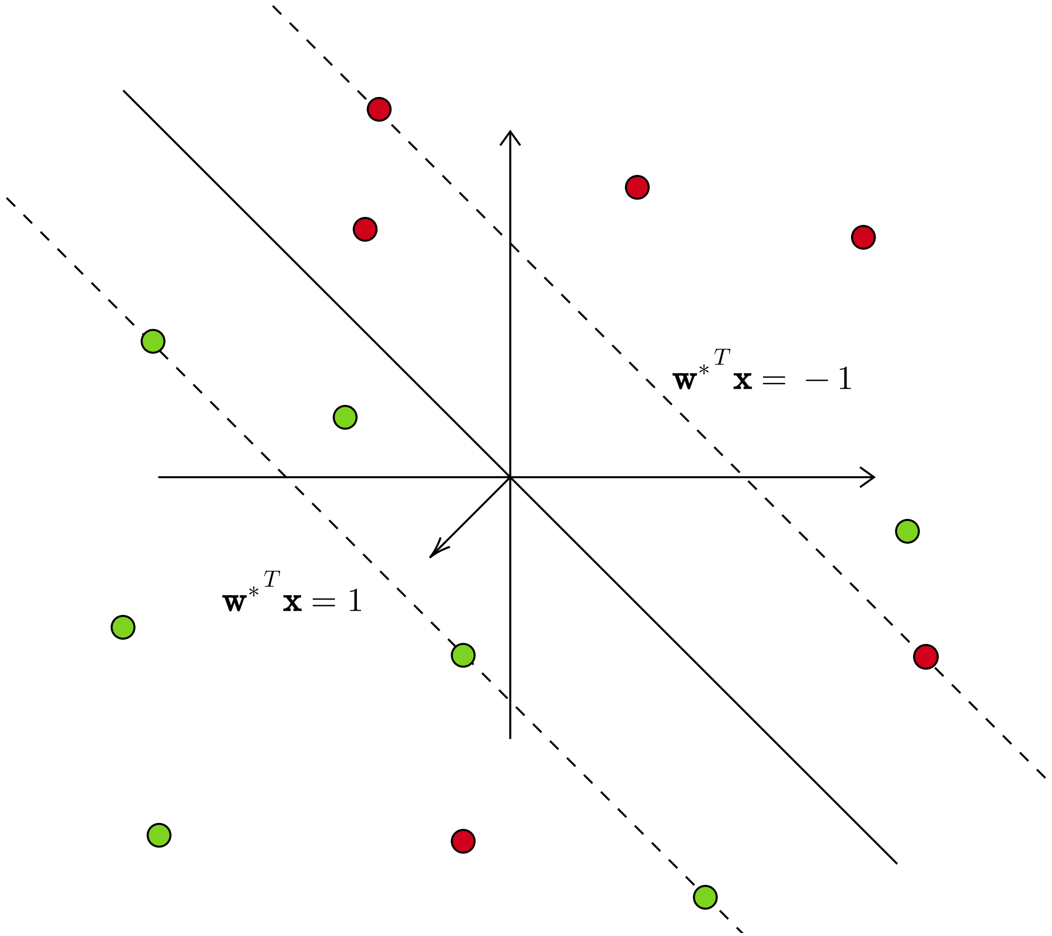
Q-12

Is this a hard-margin or soft-margin SVM?

What are the maximum number of support vectors for this problem?

Which of these points are certainly support vectors?

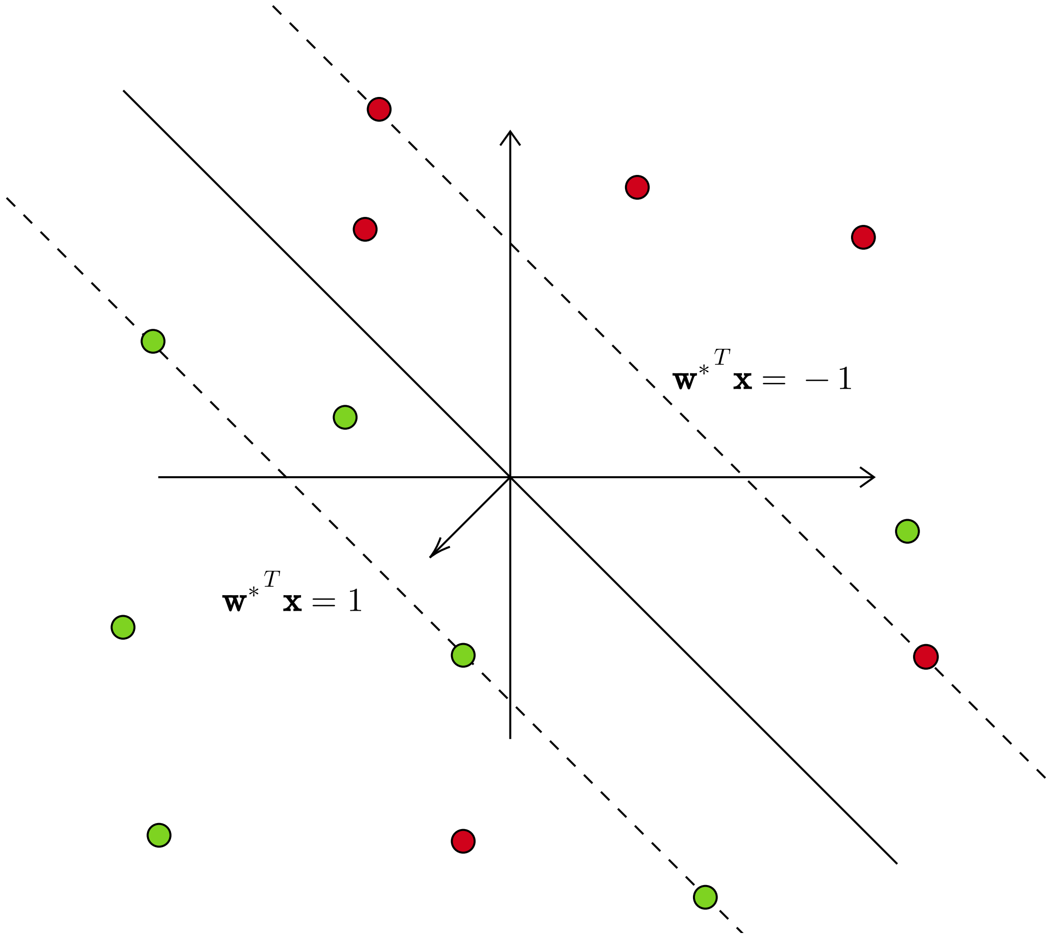
What is the value of α_i^* for these trustworthy vectors?



Q-12 Is this a hard-margin or soft-margin SVM?
What are the maximum number of support vectors for this problem?
Which of these points are certainly support vectors?
What is the value of α_i^* for these trustworthy vectors?

$$\left(\mathbf{w^*}^T \mathbf{x}_i\right) y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$

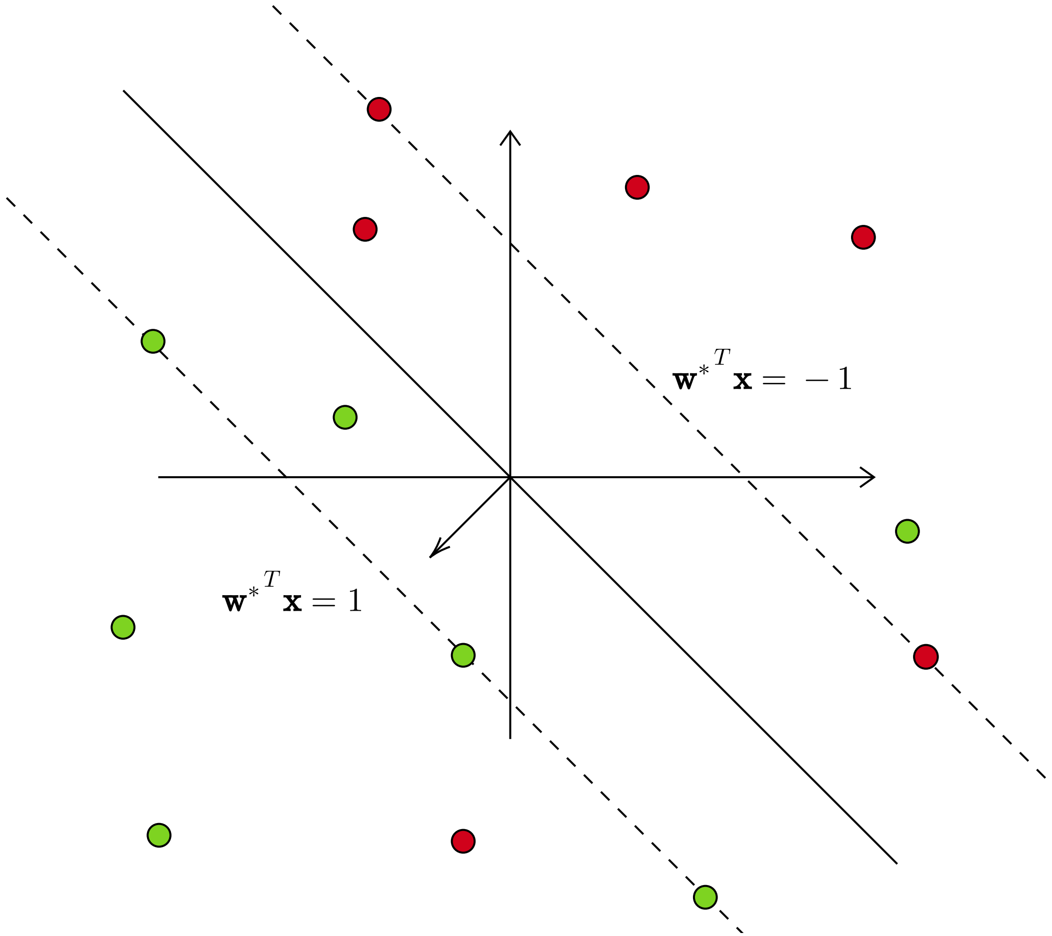


Q-12

Is this a hard-margin or soft-margin SVM?
What are the maximum number of support vectors for this problem?
Which of these points are certainly support vectors?
What is the value of α_i^* for these trustworthy vectors?

$$\left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i - \xi_i^*\right] = 0 \qquad (1)$$

Q-12 Is this a hard-margin or soft-margin SVM?

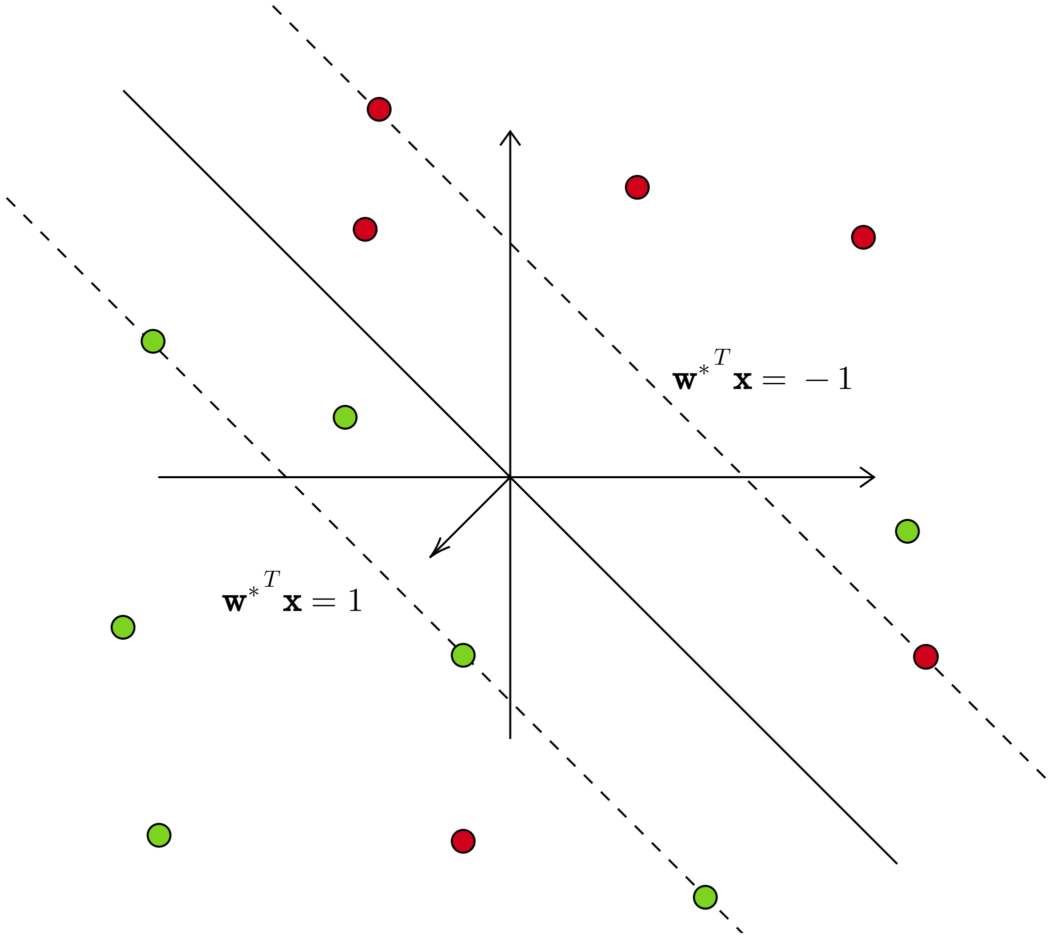
What are the maximum number of support vectors for this problem?

Which of these points are certainly support vectors?

What is the value of α_i^* for these trustworthy vectors?

$$\left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i - \xi_i^*\right] = 0 \qquad (1)$$

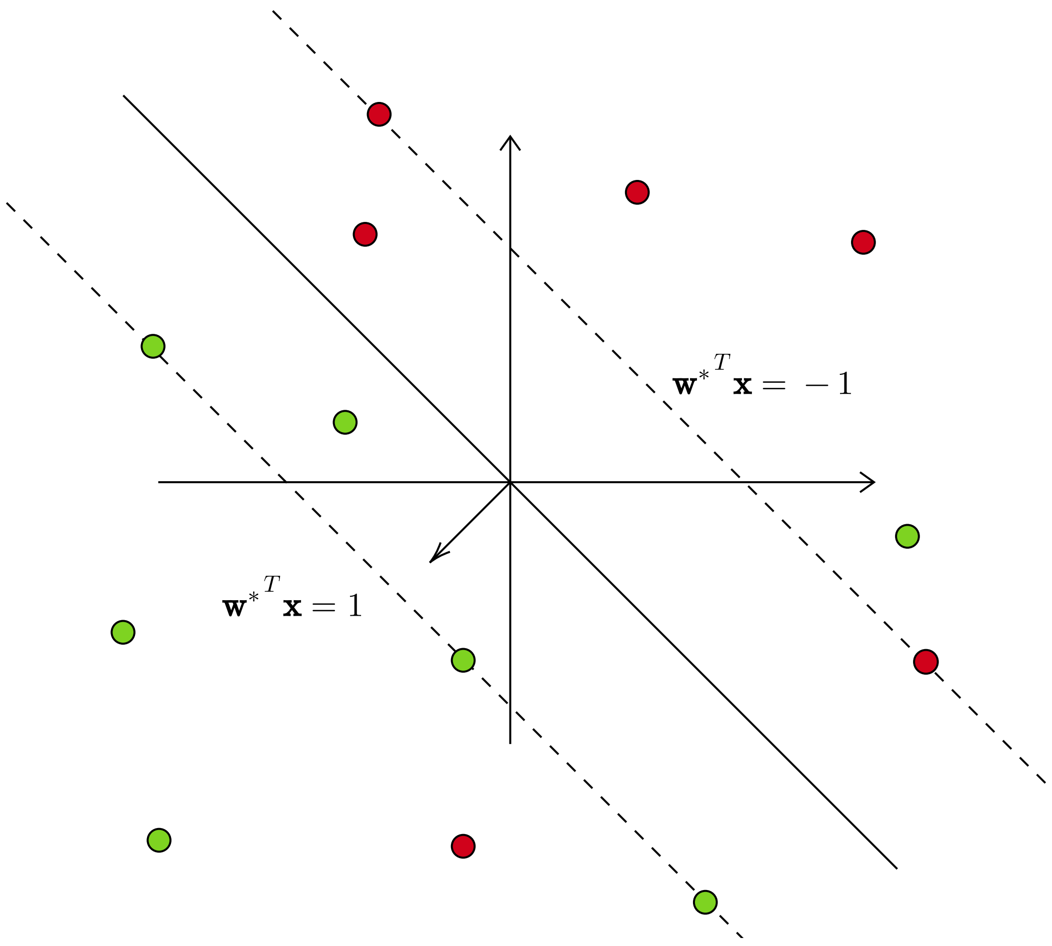
$$\beta_i^* \cdot \xi_i^* = 0 \qquad (2)$$

Q-12

Is this a hard-margin or soft-margin SVM?
What are the maximum number of support vectors for this problem?
Which of these points are certainly support vectors?
What is the value of α_i^* for these trustworthy vectors?

$$\left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i - \xi_i^*\right] = 0 \qquad (1)$$

$$\beta_i^* \cdot \xi_i^* = 0 \qquad (2)$$

$$\alpha_i^* + \beta_i^* = C \qquad (3)$$

Q-12

Is this a hard-margin or soft-margin SVM?

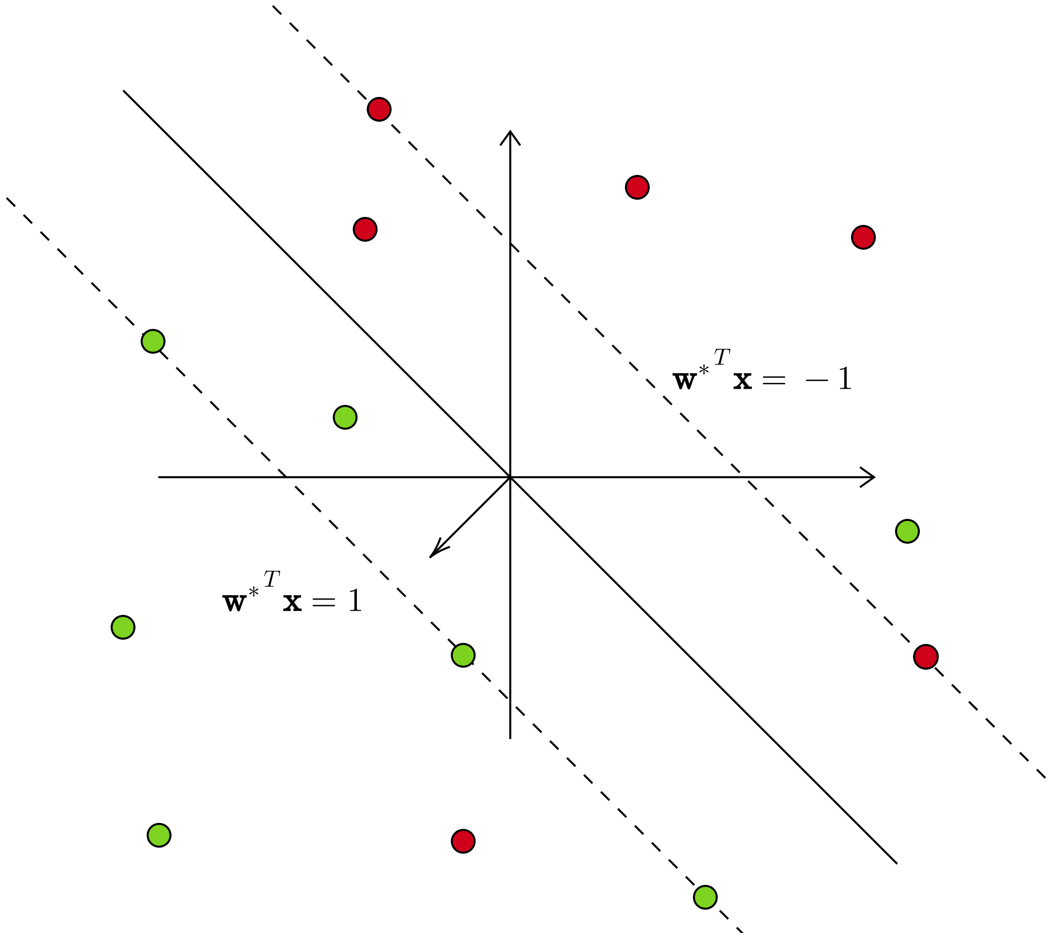
What are the maximum number of support vectors for this problem?

Which of these points are certainly support vectors?

What is the value of α_i^* for these trustworthy vectors?

$$\left({\mathbf{w}^*}^T \mathbf{x}_i\right) y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left({\mathbf{w}^*}^T \mathbf{x}_i\right) y_i - \xi_i^*\right] = 0 \qquad (1)$$

$$\beta_i^* \cdot \xi_i^* = 0 \qquad (2)$$

$$\alpha_i^* + \beta_i^* = C \qquad (3)$$

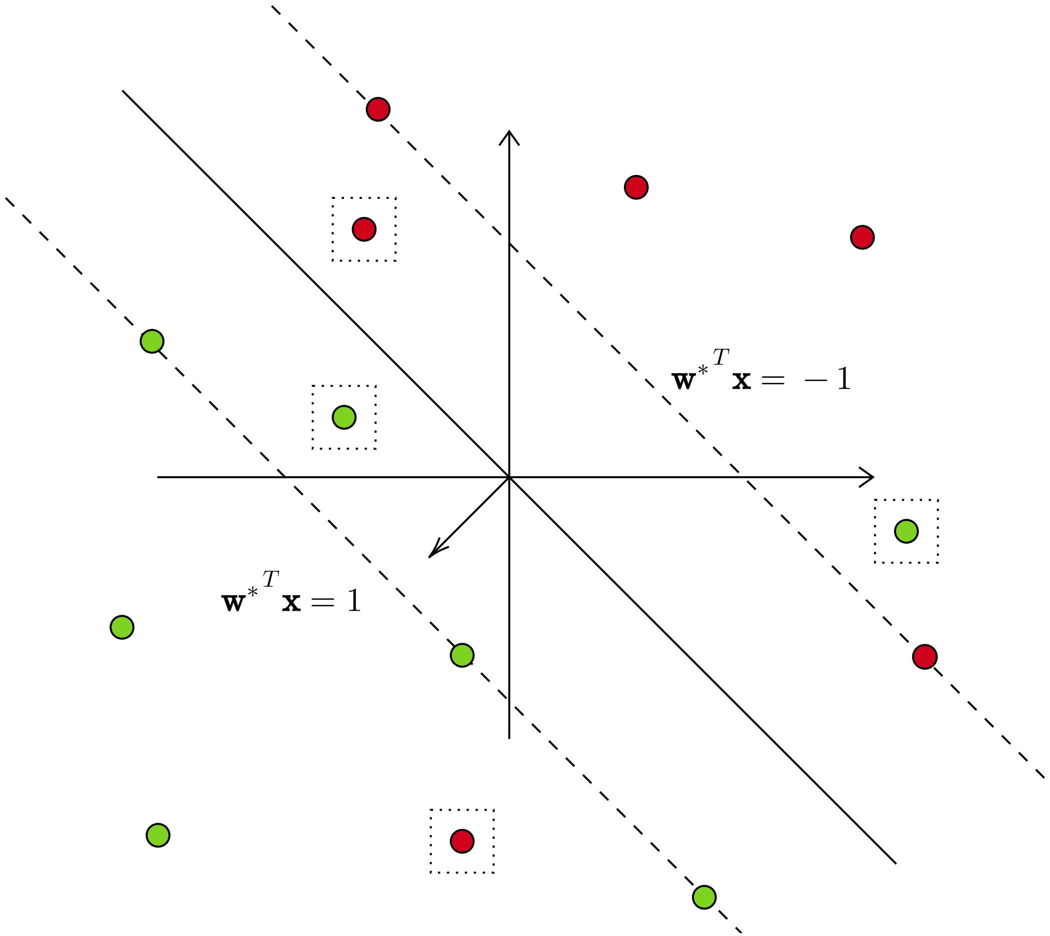
$$\leqslant n\left(\mathsf{SV}\right) \leqslant$$

Q-12

Is this a hard-margin or soft-margin SVM?
What are the maximum number of support vectors for this problem?
Which of these points are certainly support vectors?
What is the value of α_i^* for these trustworthy vectors?

$$\left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left(\mathbf{w}^{*T}\mathbf{x}_i\right)y_i - \xi_i^*\right] = 0 \qquad (1)$$

$$\beta_i^* \cdot \xi_i^* = 0 \qquad (2)$$

$$\alpha_i^* + \beta_i^* = C \qquad (3)$$

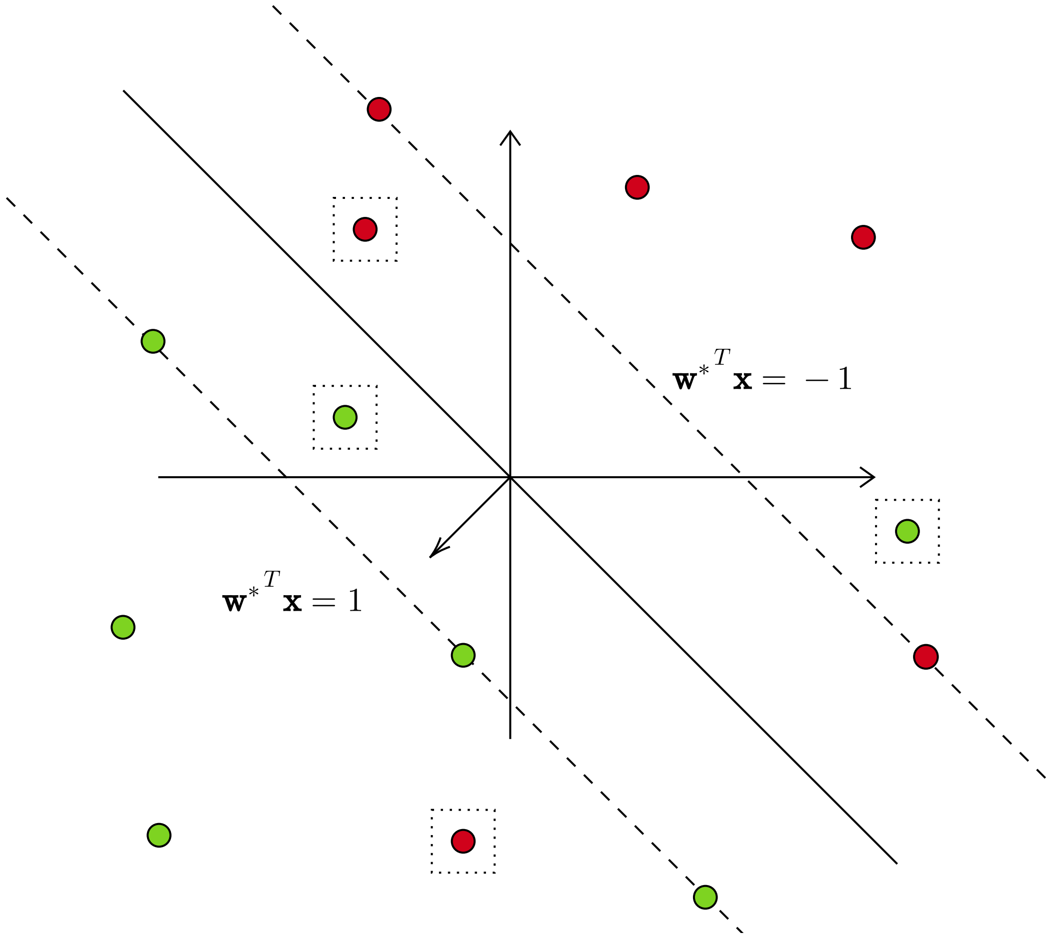
$$4 \leqslant n(\mathsf{SV}) \leqslant$$

Q-12

Is this a hard-margin or soft-margin SVM?
What are the maximum number of support vectors for this problem?
Which of these points are certainly support vectors?
What is the value of α_i^* for these trustworthy vectors?

$$\left({\mathbf{w}^*}^T \mathbf{x}_i\right) y_i + \xi_i^* \geqslant 1 \qquad (1)$$

$$\xi_i^* \geqslant 0 \qquad (2)$$



$$\alpha_i^* \cdot \left[1 - \left({\mathbf{w}^*}^T \mathbf{x}_i\right) y_i - \xi_i^*\right] = 0 \qquad (1)$$

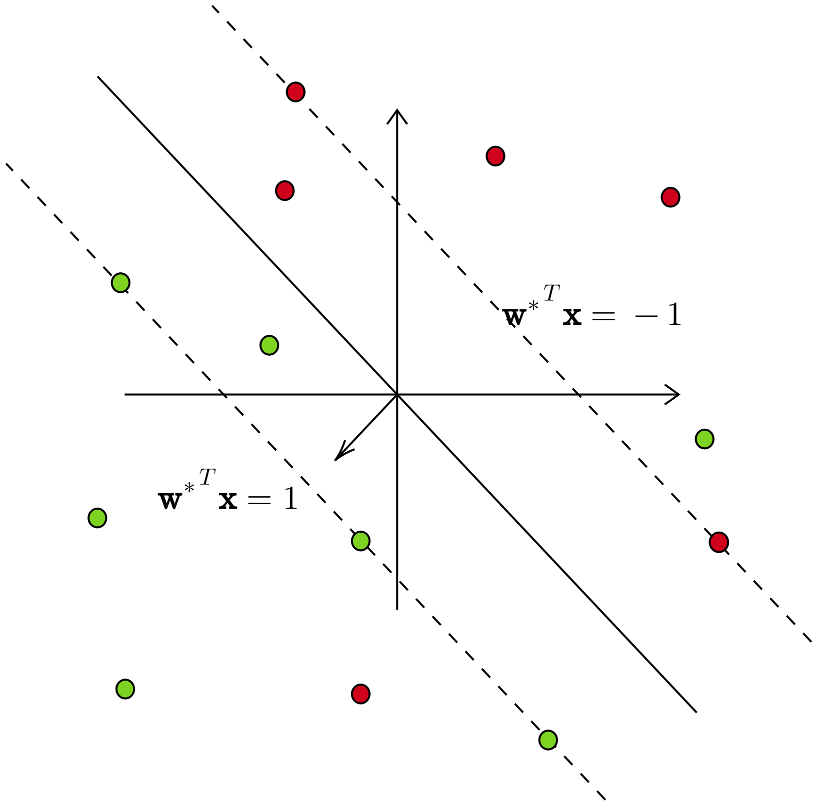
$$\beta_i^* \cdot \xi_i^* = 0 \qquad (2)$$

$$\alpha_i^* + \beta_i^* = C \qquad (3)$$

$$4 \leqslant n(\mathsf{SV}) \leqslant 9$$

Q-12

Since we have the image of the feature space given to us, it would be worthwhile to look at the **primal picture**. We have three cases:



Beyond the margin

$$\boxed{\left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i > 1} \longrightarrow \xi_i^* = 0$$

$$\boxed{\left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i > 1} \longrightarrow \left[1 - \left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i - \xi_i^*\right] \neq 0 \longrightarrow \boxed{\alpha_i^* = 0}$$

On the margin

$$\boxed{\left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i = 1} \longrightarrow \xi_i^* = 0$$

$$\boxed{\left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i = 1} \longrightarrow \boxed{0 \leqslant \alpha_i^* \leqslant C}$$

Margin violation

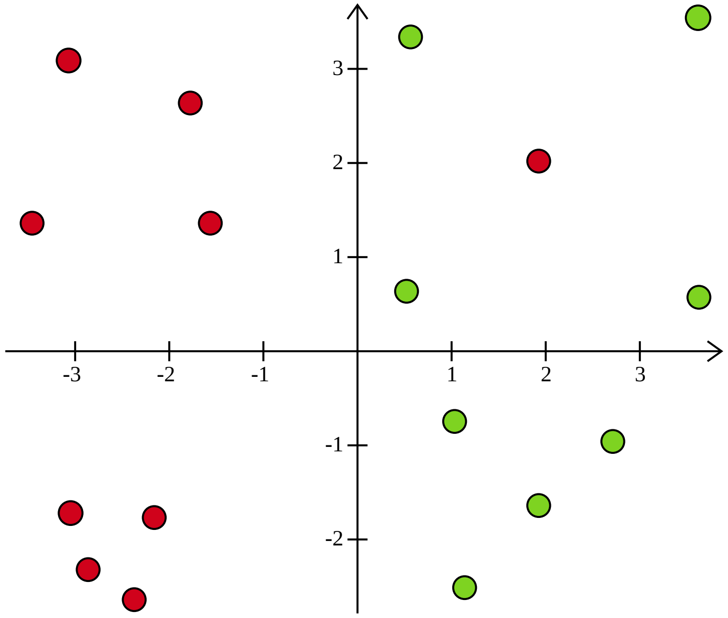
$$\boxed{\left(\mathbf{w^*}^T\mathbf{x}_i\right)y_i < 1} \longrightarrow \xi_i^* > 0 \longrightarrow \beta_i^* = 0 \longrightarrow \boxed{\alpha_i^* = C}$$

Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.



Q-13

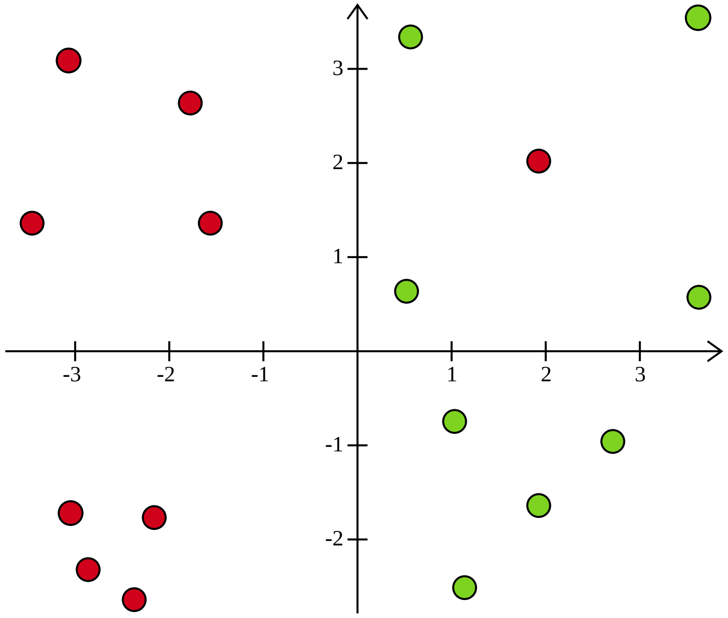
(a) Deep trees perform well on the training data.

(b) Deep trees perform well on both training and test data.

(c) Deep trees perform well on the training data but do not perform well on the test data.

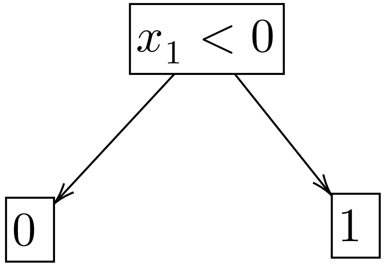
(d) Deep trees perform poorly on both training and test data.

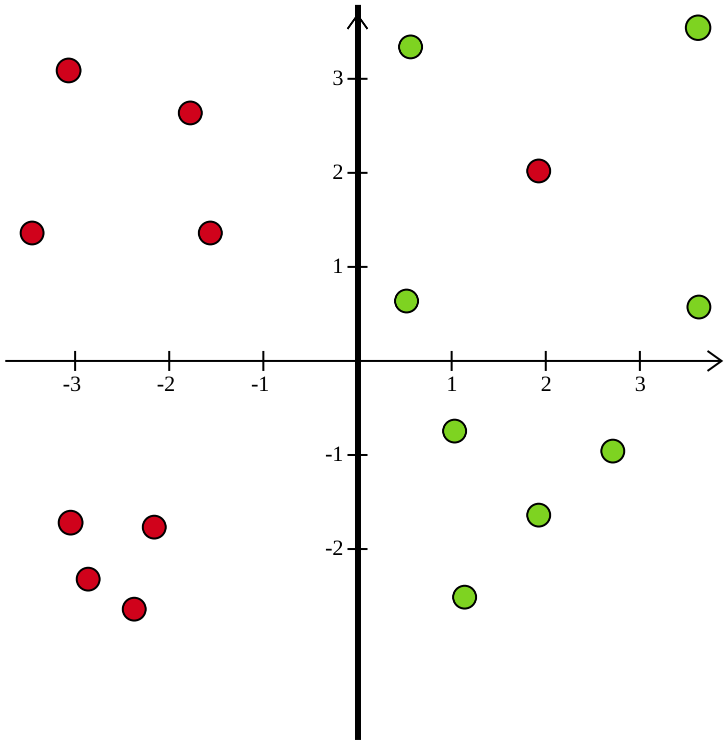
$\boxed{x_1 < 0}$



Q-13

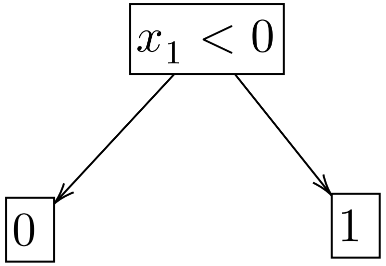
(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.



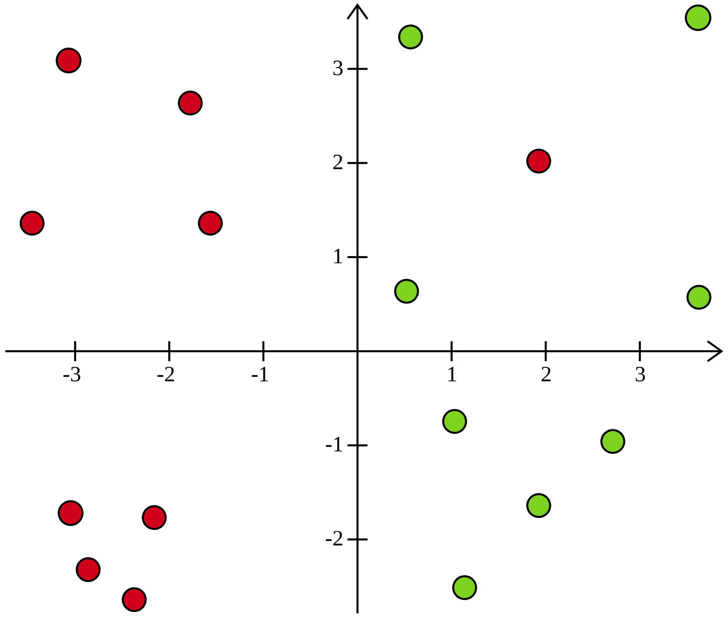


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

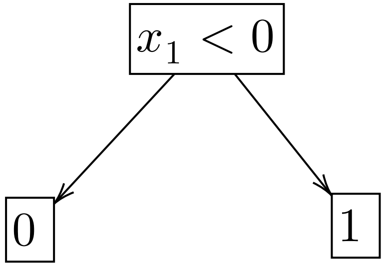


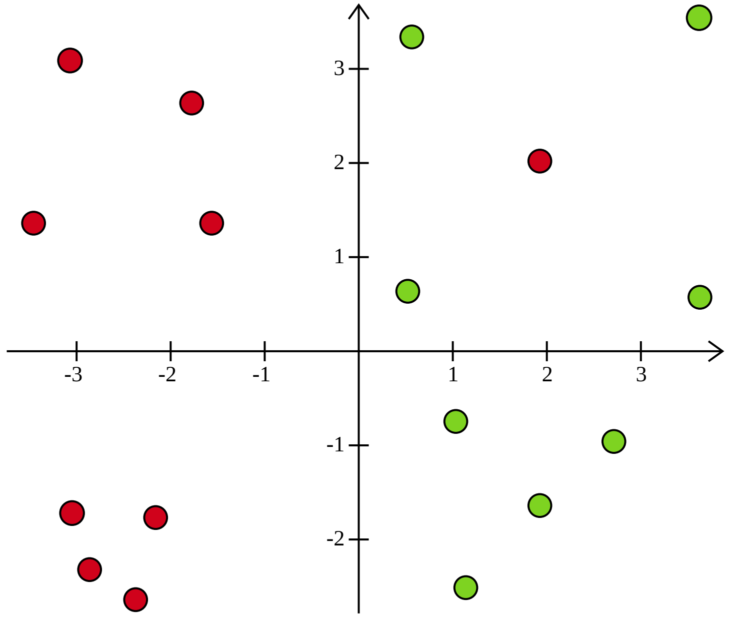
$\boxed{x_1 < 0}$

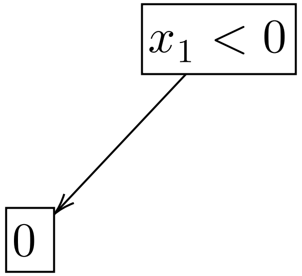


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

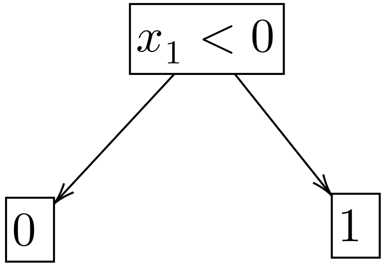


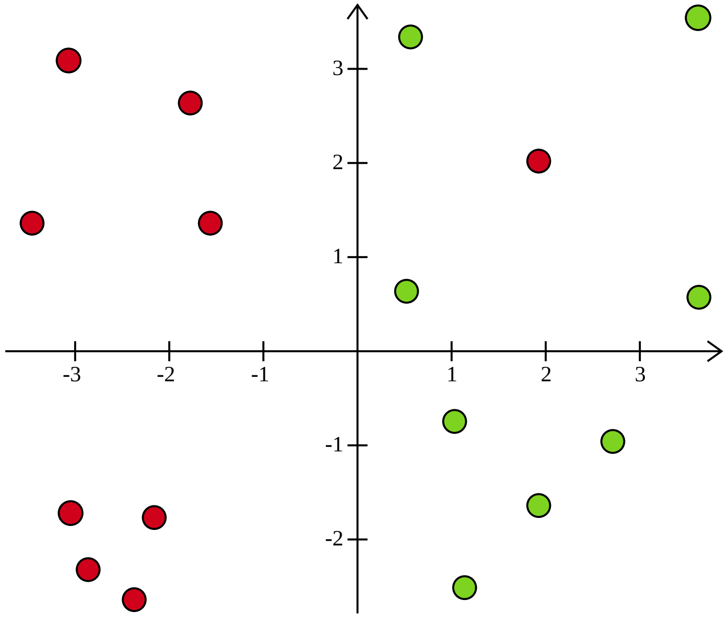


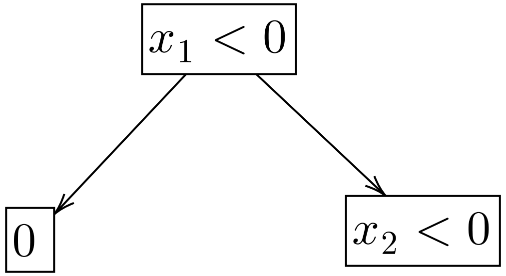


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

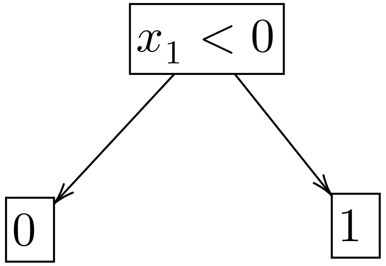


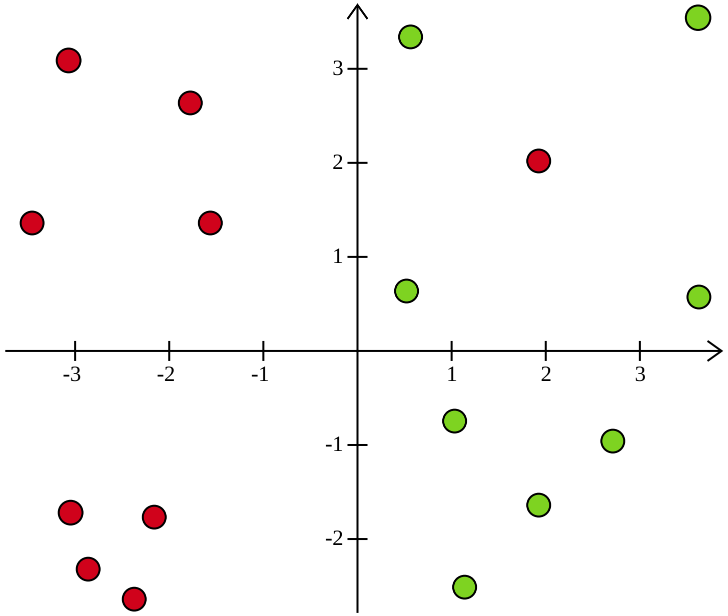


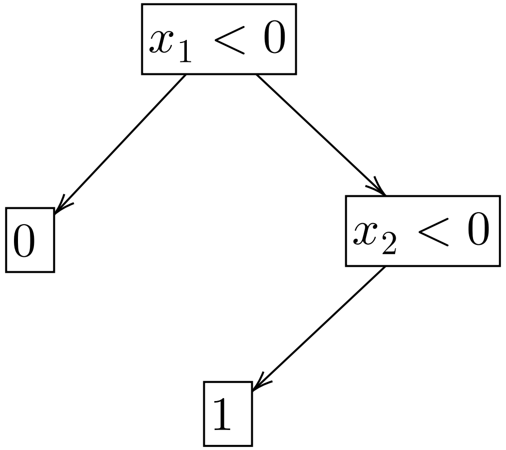


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

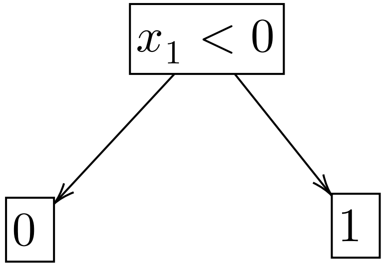


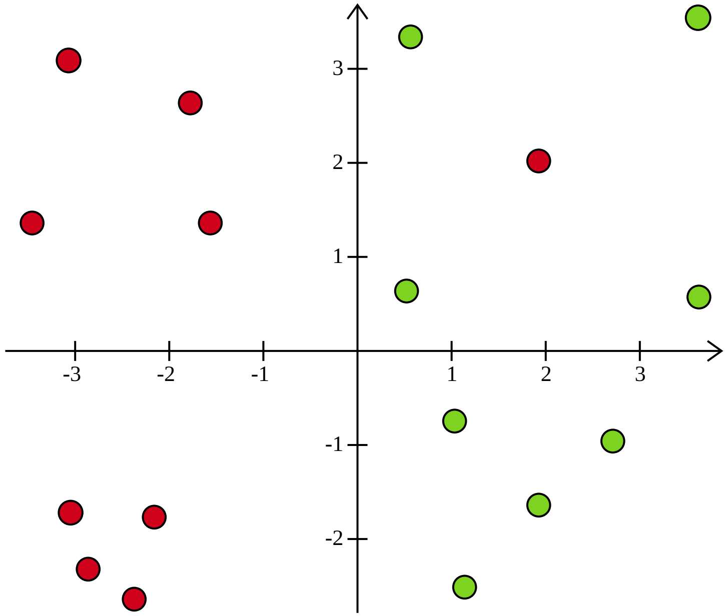


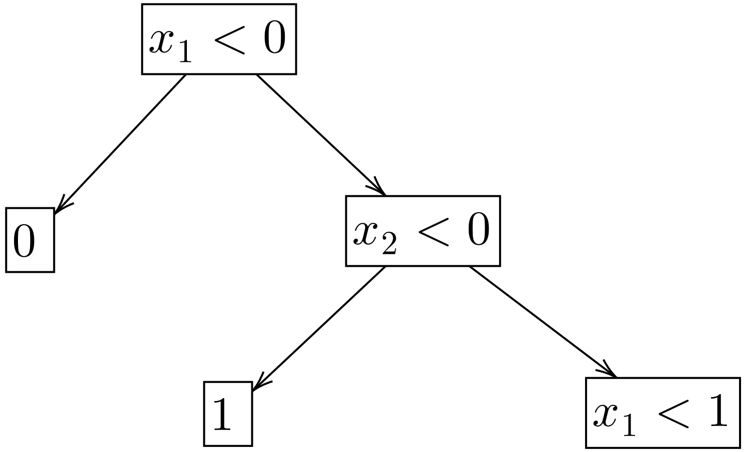


Q-13

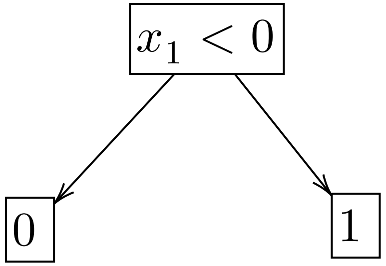
(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

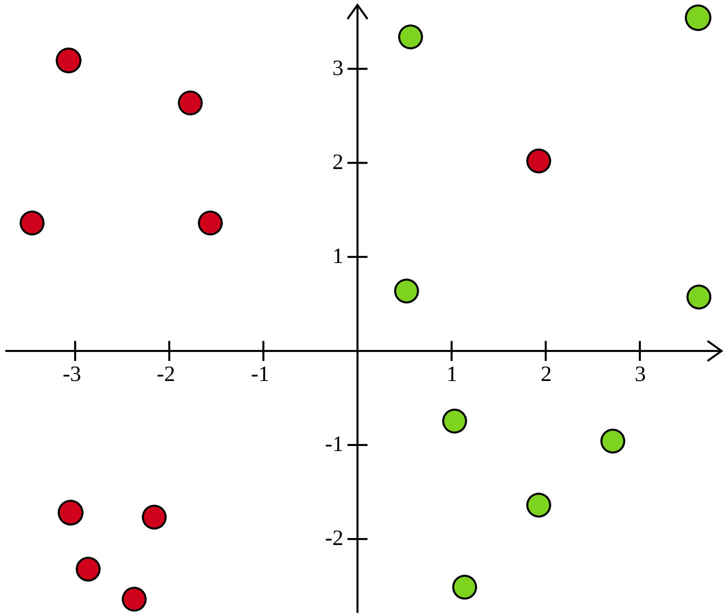


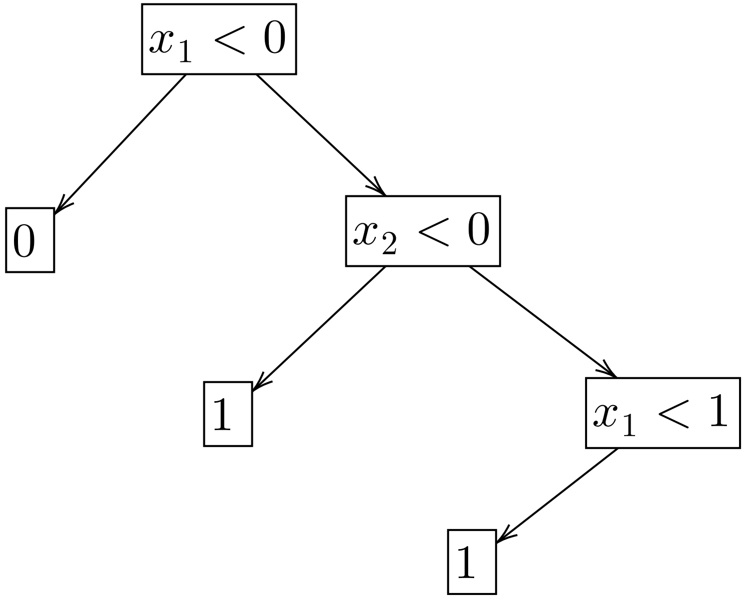




Q-13 (a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

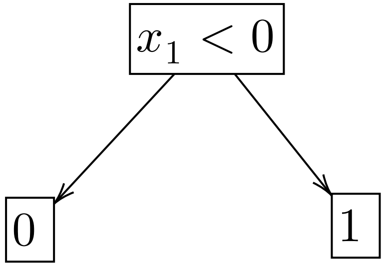


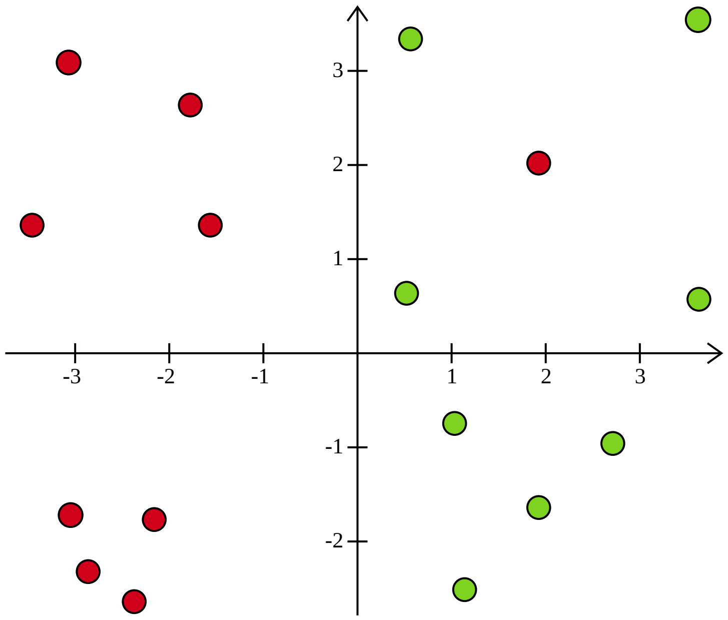


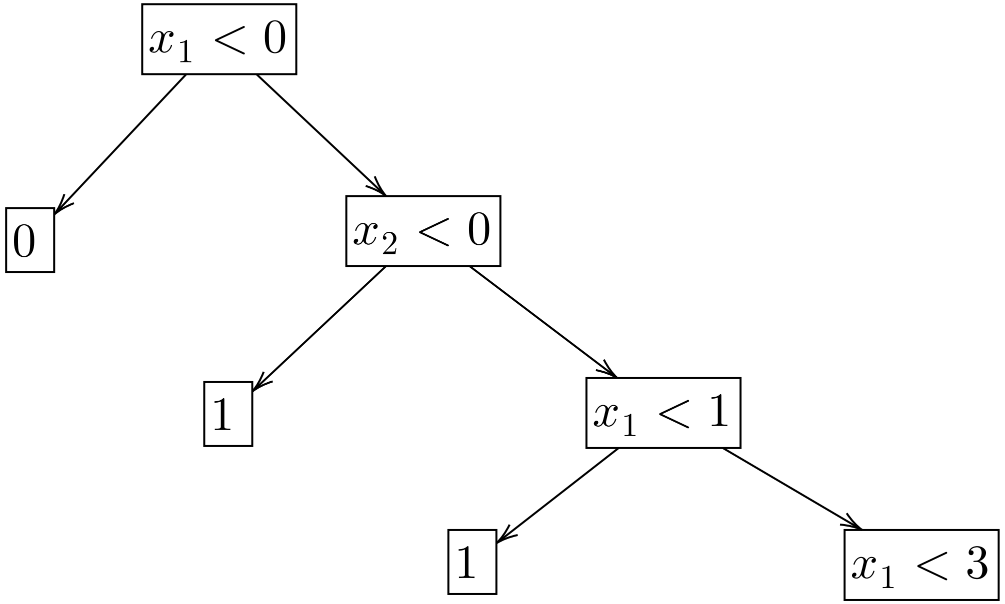


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

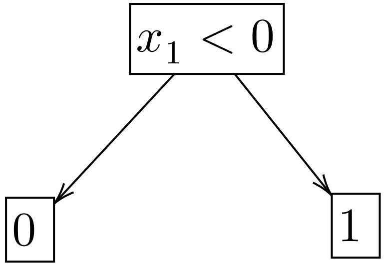


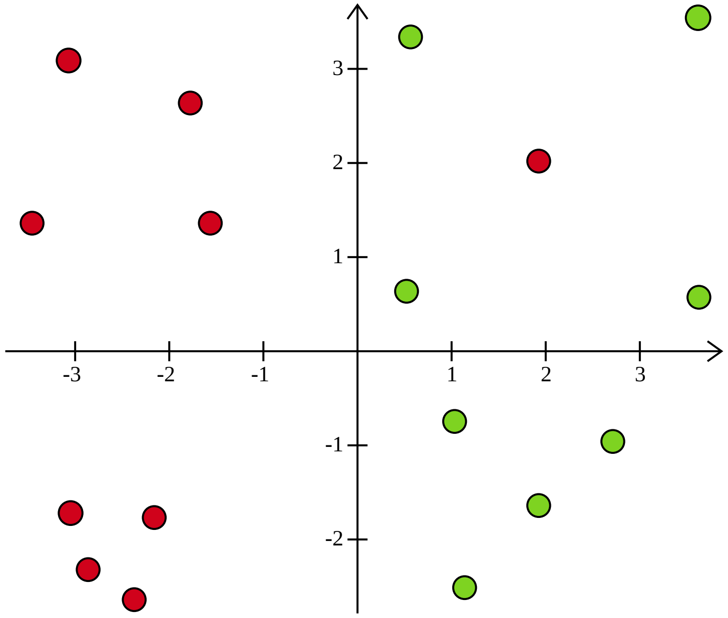


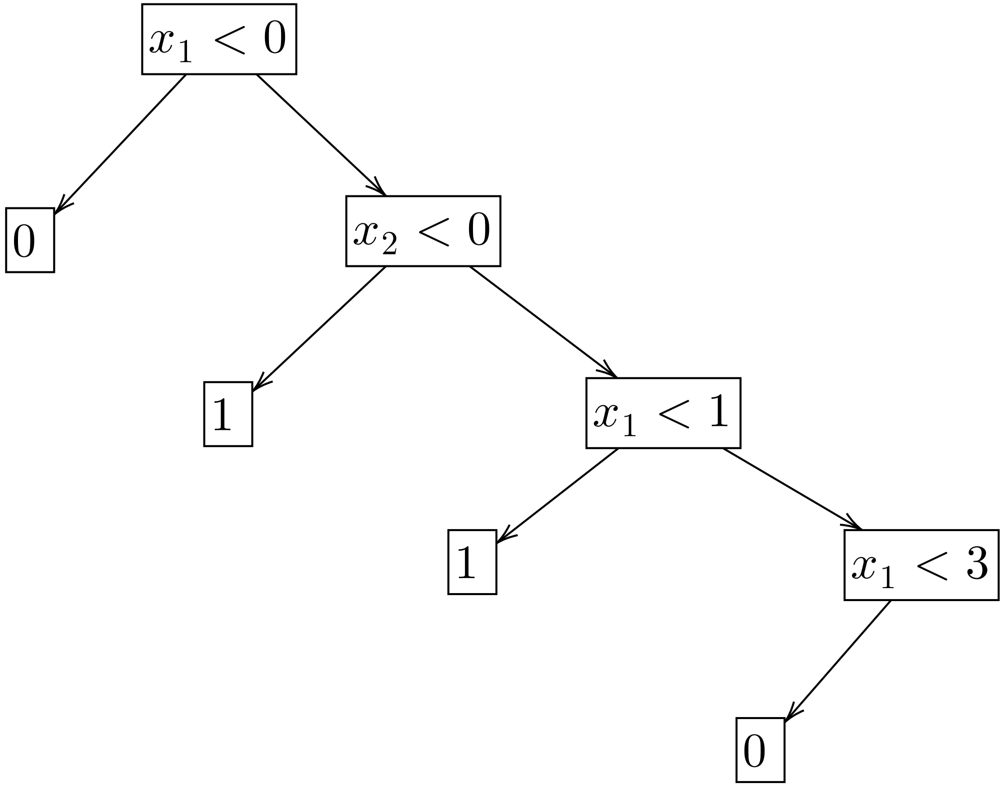


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

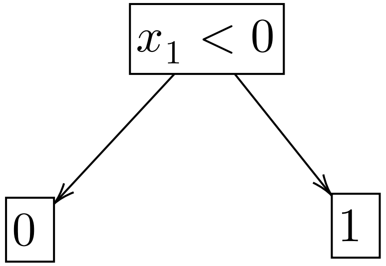


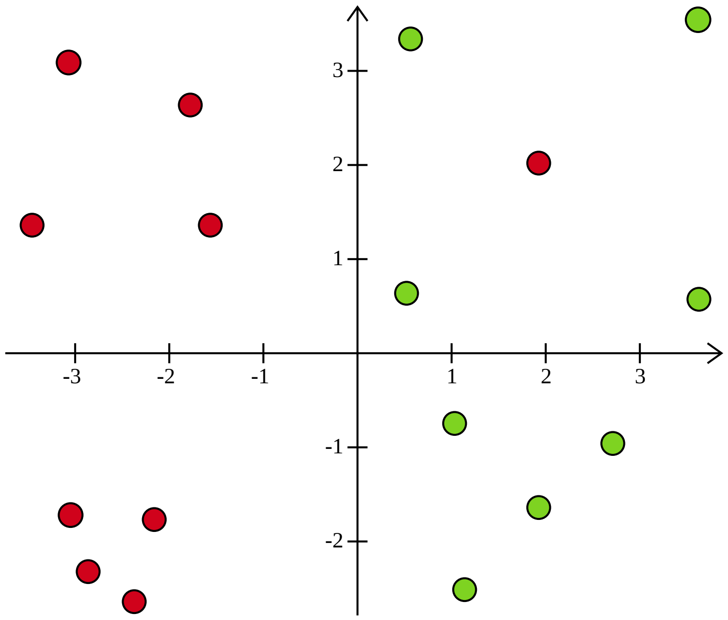


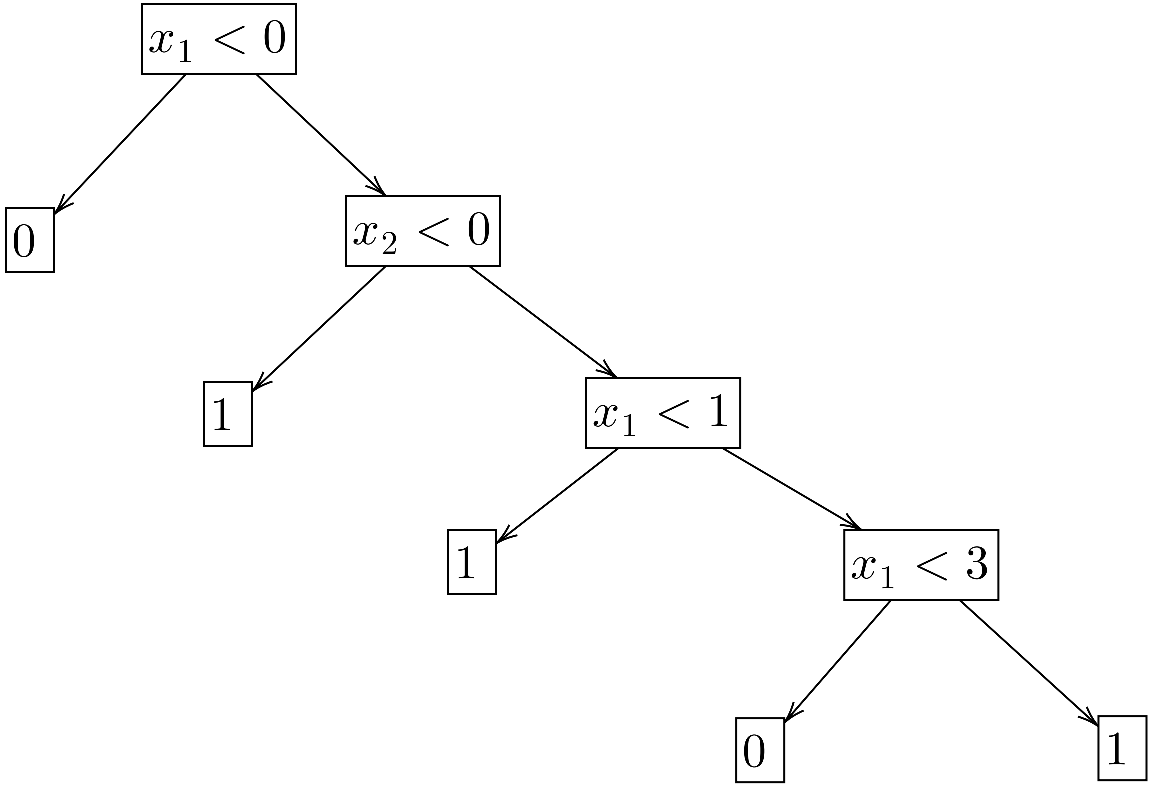


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.

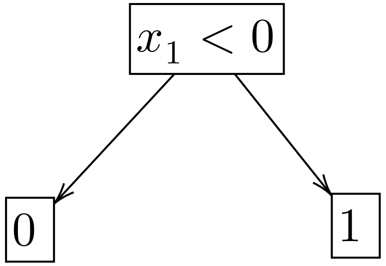


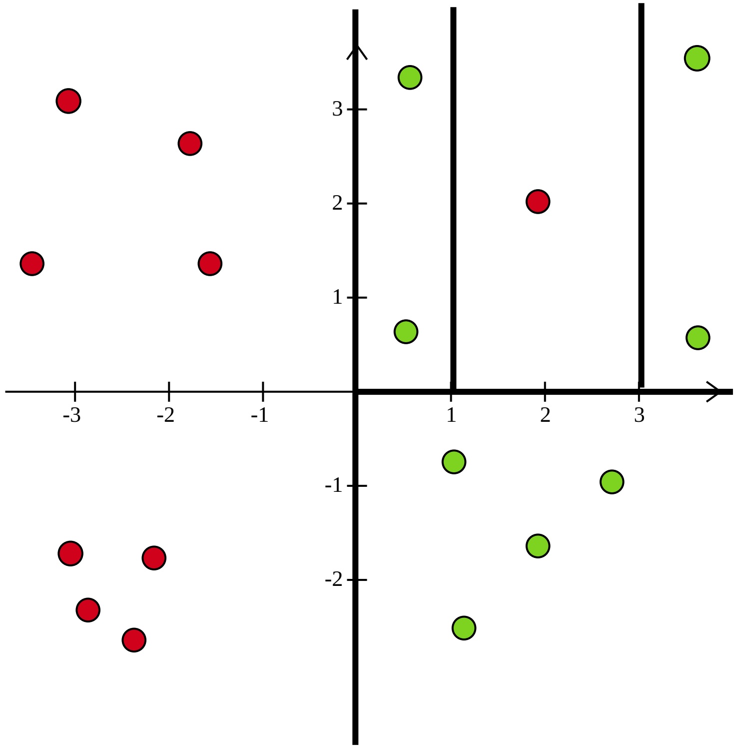


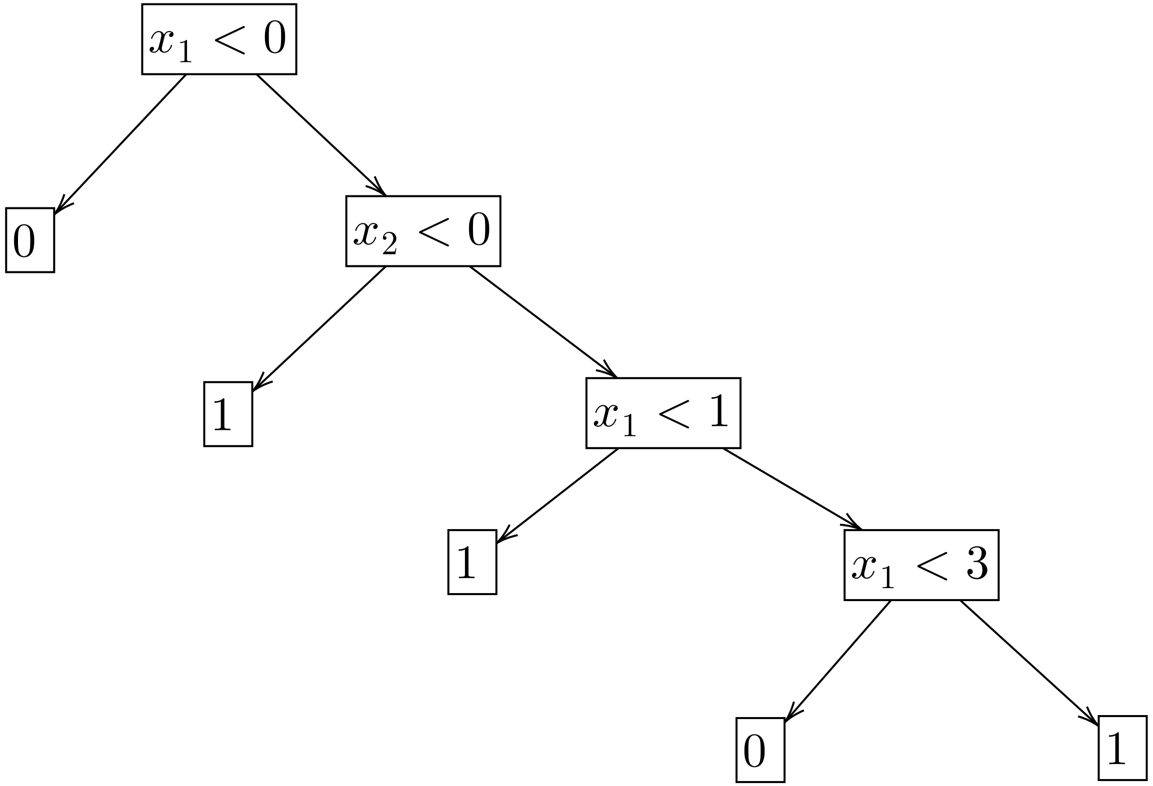


Q-13

(a) Deep trees perform well on the training data.
(b) Deep trees perform well on both training and test data.
(c) Deep trees perform well on the training data but do not perform well on the test data.
(d) Deep trees perform poorly on both training and test data.







Q-14

- Vector output by some hidden layer in a neural network:
  $[0.1 \quad 0.8 \quad 0. \quad 0.5 \quad 0.7 \quad 0.9]^T$
- How many neurons does this layer have?
- What is the activation function used in this layer?
- How many neurons does the next layer have?

Q-14

- Vector output by some hidden layer in a neural network:
  $[0.1 \quad 0.8 \quad 0. \quad 0.5 \quad 0.7 \quad 0.9]^T$
- How many neurons does this layer have?
- What is the activation function used in this layer?
- How many neurons does the next layer have?

h_1 0.1

h_2 0.8

h_3 0

h_4 0.5

h_5 0.7

h_6 0.9

Q-14

- `Vector output by some hidden layer in a neural network:`
  $[0.1 \quad 0.8 \quad 0. \quad 0.5 \quad 0.7 \quad 0.9]^T$
- `How many neurons does this layer have?`
- `What is the activation function used in this layer?`
- `How many neurons does the next layer have?`

h_1 0.1

h_2 0.8

h_3 0

h_4 0.5

h_5 0.7

h_6 0.9

$\texttt{ReLU}(z) = \max(0, z)$

$\texttt{ReLU}(z) \geqslant 0$

Q-14

- Vector output by some hidden layer in a neural network:
  $[0.1 \quad 0.8 \quad 0. \quad 0.5 \quad 0.7 \quad 0.9]^T$
- How many neurons does this layer have?
- What is the activation function used in this layer?
- How many neurons does the next layer have?

h_1	0.1
h_2	0.8
h_3	0
h_4	0.5
h_5	0.7
h_6	0.9

$\texttt{ReLU}(z) = \max(0, z)$ $\texttt{ReLU}(z) \geqslant 0$

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$ $0 < \sigma(z) < 1$

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$u =$

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$		
Squared loss		
Hinge loss		
Logisitc loss		
Modified hinge loss		

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	
Squared loss		
Hinge loss		
Logisitc loss		
Modified hinge loss		

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T\mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	
Squared loss	$(u-1)^2$	
Hinge loss		
Logisitc loss		
Modified hinge loss		

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	
Squared loss	$(u-1)^2$	
Hinge loss	$\max(0, 1-u)$	
Logisitc loss		
Modified hinge loss		

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	
Squared loss	$(u-1)^2$	
Hinge loss	$\max(0, 1-u)$	
Logisitc loss	$\ln\left(1+e^{-u}\right)$	
Modified hinge loss		

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	
Squared loss	$(u-1)^2$	
Hinge loss	$\max(0, 1-u)$	
Logisitc loss	$\ln\left(1+e^{-u}\right)$	
Modified hinge loss	$\max(0, -u)$	

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T\mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	
Hinge loss	$\max(0, 1-u)$	
Logisitc loss	$\ln\left(1+e^{-u}\right)$	
Modified hinge loss	$\max(0, -u)$	

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	4
Hinge loss	$\max(0, 1-u)$	
Logisitc loss	$\ln\left(1+e^{-u}\right)$	
Modified hinge loss	$\max(0, -u)$	

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	4
Hinge loss	$\max(0, 1-u)$	2
Logisitc loss	$\ln\left(1+e^{-u}\right)$	
Modified hinge loss	$\max(0, -u)$	

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	4
Hinge loss	$\max(0, 1-u)$	2
Logisitc loss	$\ln\left(1+e^{-u}\right)$	1.313
Modified hinge loss	$\max(0, -u)$	

Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

$$u = \left(\mathbf{w}^T \mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	4
Hinge loss	$\max(0, 1-u)$	2
Logisitc loss	$\ln\left(1+e^{-u}\right)$	1.313
Modified hinge loss	$\max(0, -u)$	1

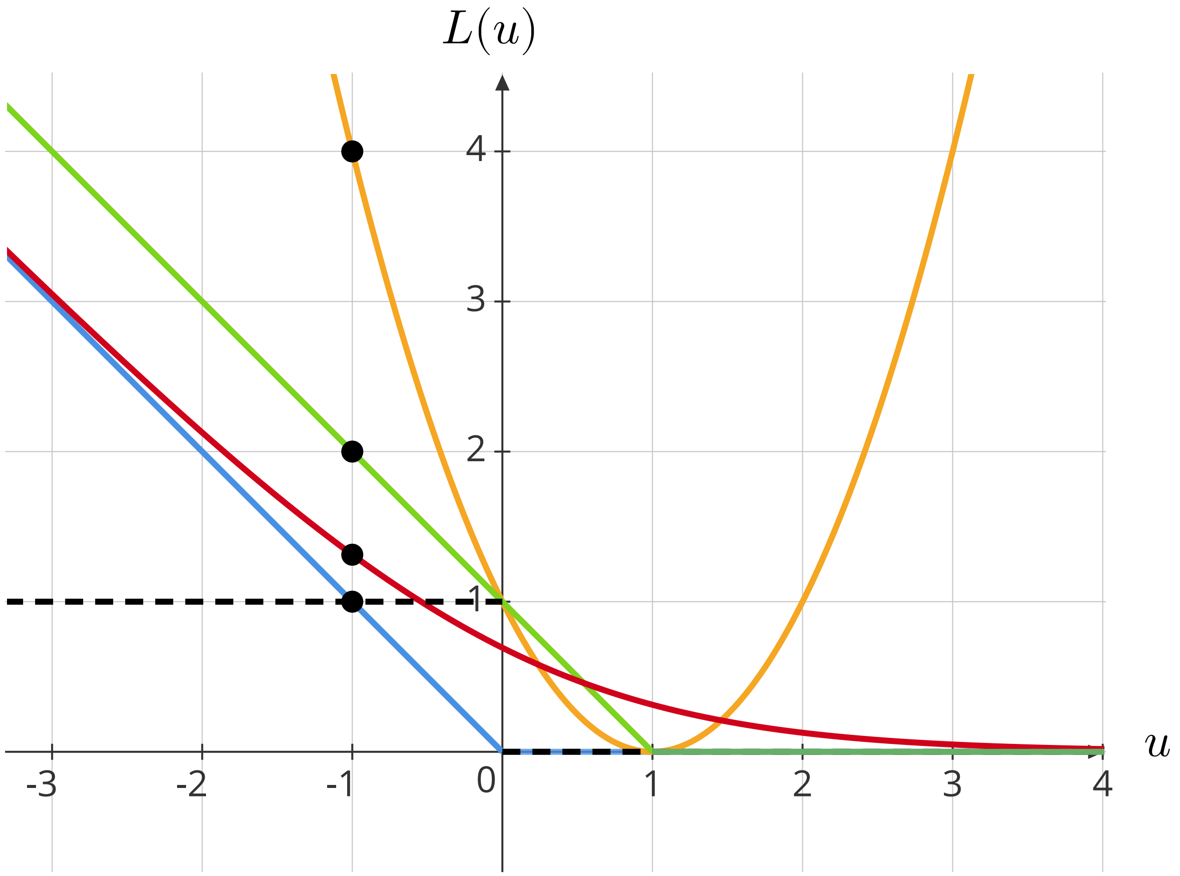
Q-15

Compute the loss for a linear classifier with $\mathbf{w} = [1 \quad 0 \quad -1]^T$ and $\mathbf{x} = [2 \quad 3 \quad 1]^T, y = -1$

- 0-1 loss
- Squared loss
- Hinge loss
- Logisitc loss
- Modified hinge loss

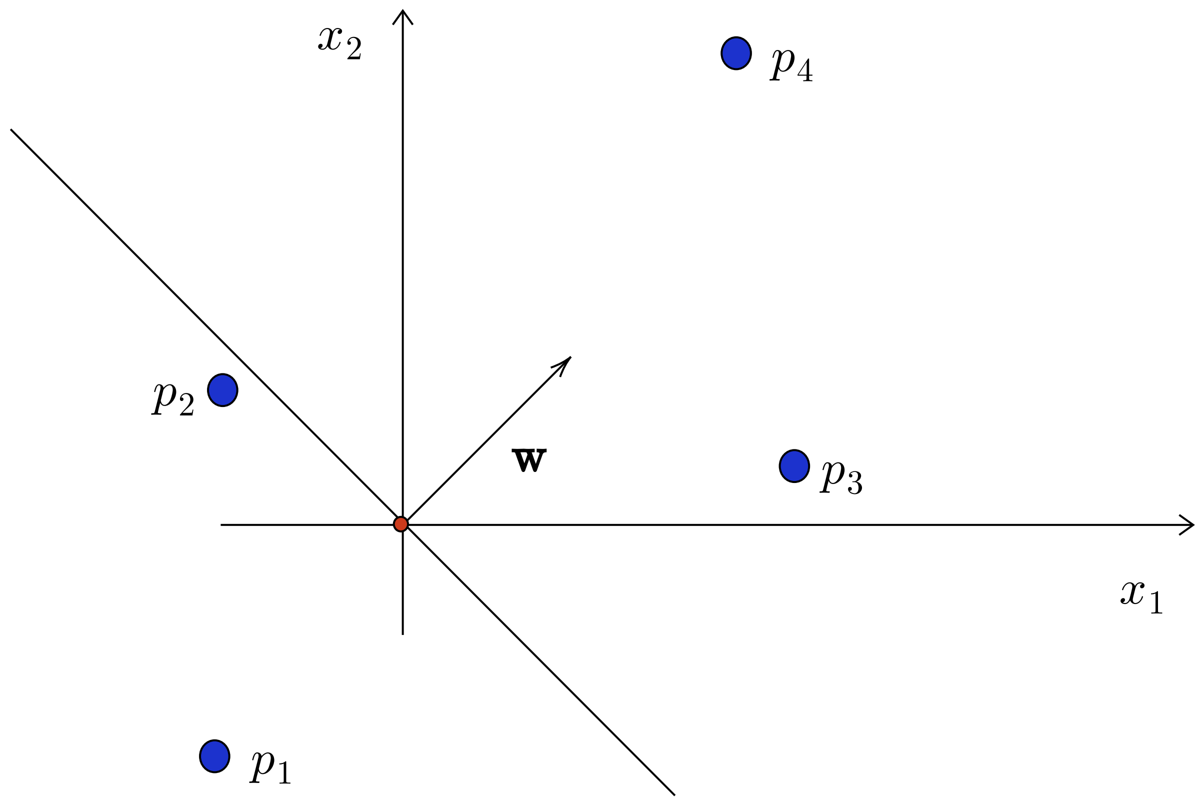
$$u = \left(\mathbf{w}^T\mathbf{x}\right) \cdot y = -1$$

Loss	$L(u)$	$L(-1)$
$0-1$	$\begin{cases} 0, & u \geqslant 0 \\ 1, & u < 0 \end{cases}$	1
Squared loss	$(u-1)^2$	4
Hinge loss	$\max(0, 1-u)$	2
Logisitc loss	$\ln\left(1+e^{-u}\right)$	1.313
Modified hinge loss	$\max(0, -u)$	1



Q-16

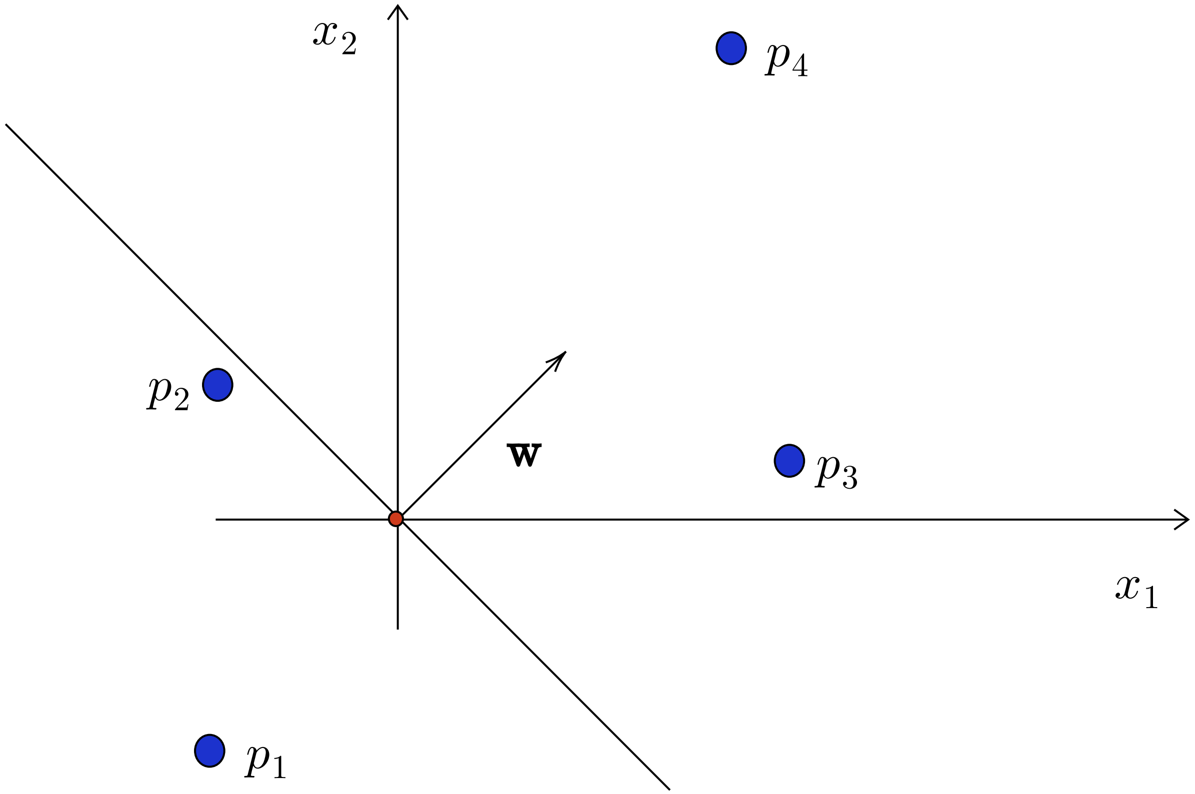
- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions



Q-16

- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions

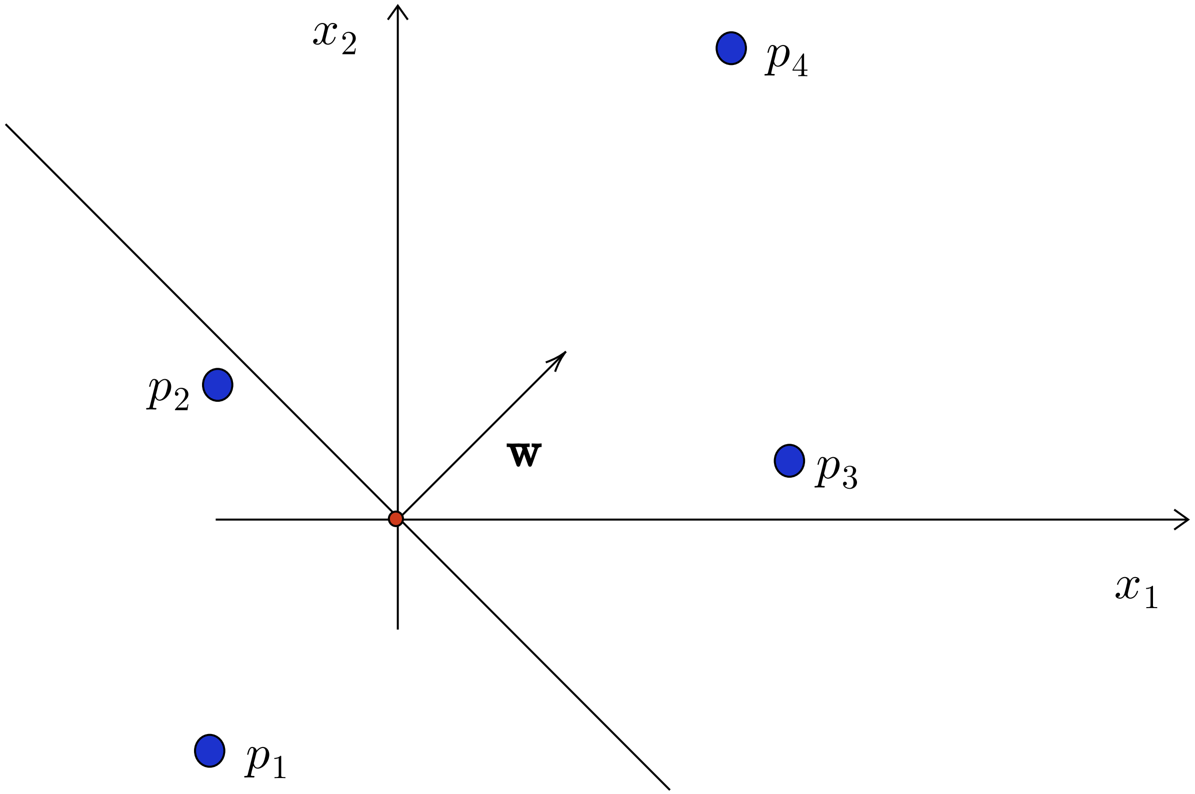
$$p_i = P(y = 1 \mid \mathbf{x}_i) = \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}}$$



Q-16

- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions

$$p_i = P(y = 1 \mid \mathbf{x}_i) = \sigma\left(\mathbf{w}^T\mathbf{x}_i\right) = \frac{1}{1 + e^{-\mathbf{w}^T\mathbf{x}_i}}$$

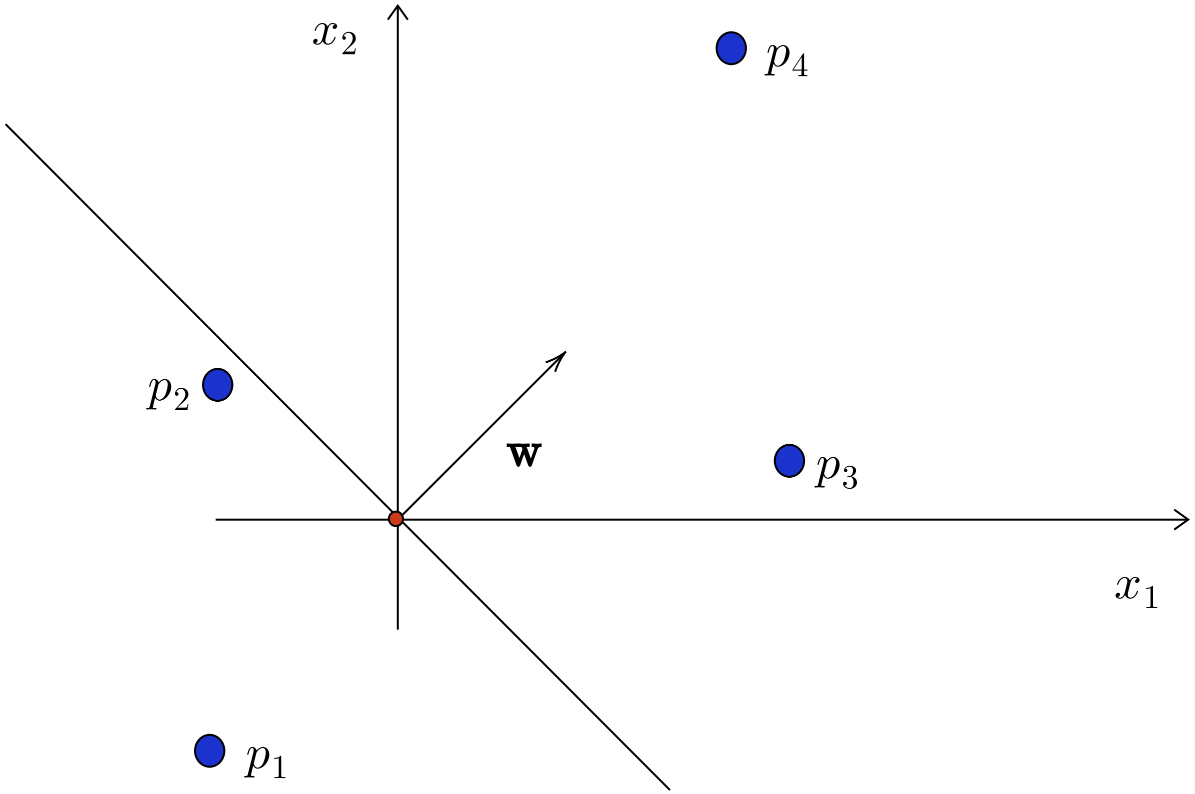


$$p_1 < p_2 < p_3 < p_4$$

Q-16

- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions

$$p_i = P(y = 1 \mid \mathbf{x}_i) = \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}}$$



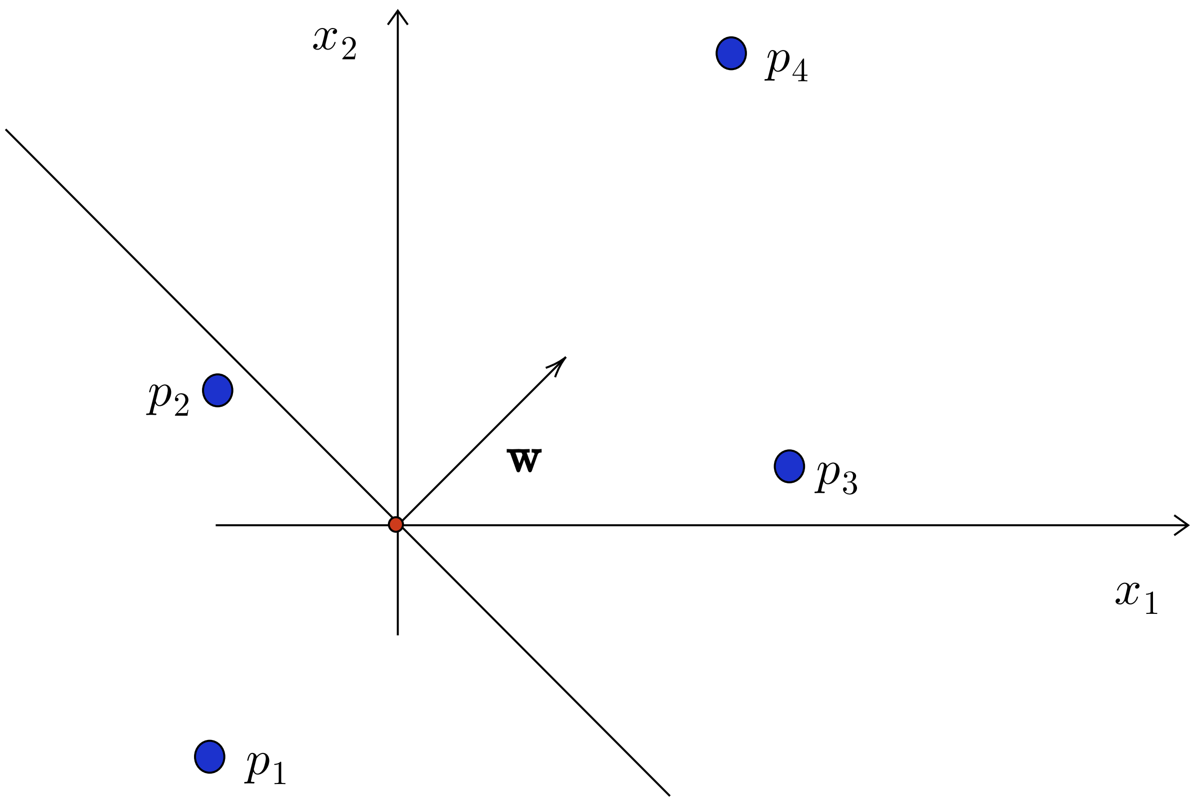
$$p_1 < p_2 < p_3 < p_4$$

$$\widehat{y} = \begin{cases} 1, & \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) \geqslant T \\ 0, & \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) < T \end{cases}$$

Q-16

- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions

$$p_i = P(y = 1 \mid \mathbf{x}_i) = \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}_i}}$$



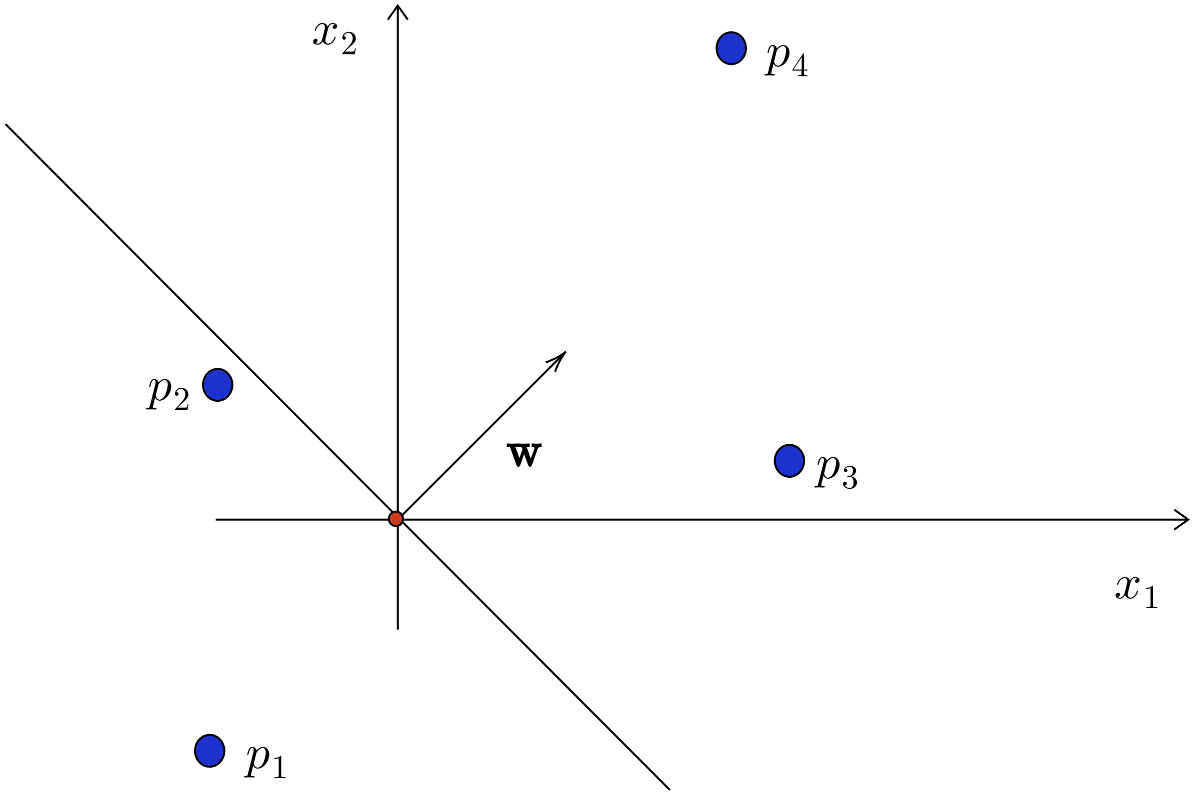
$$p_1 < p_2 < p_3 < p_4$$

$$\widehat{y} = \begin{cases} 1, & \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) \geqslant T \\ 0, & \sigma\left(\mathbf{w}^T \mathbf{x}_i\right) < T \end{cases} \qquad \widehat{y} = \begin{cases} 1, & \mathbf{w}^T \mathbf{x}_i \geqslant 0 \\ 0, & \mathbf{w}^T \mathbf{x}_i < 0 \end{cases}$$

Q-16

- Logistic regression model
- Threshold is 0.5
- Order the probabilities
- Find the predictions

$$p_i = P(y = 1 \mid \mathbf{x}_i) = \sigma\left(\mathbf{w}^T\mathbf{x}_i\right) = \frac{1}{1 + e^{-\mathbf{w}^T\mathbf{x}_i}}$$



$$p_1 < p_2 < p_3 < p_4$$

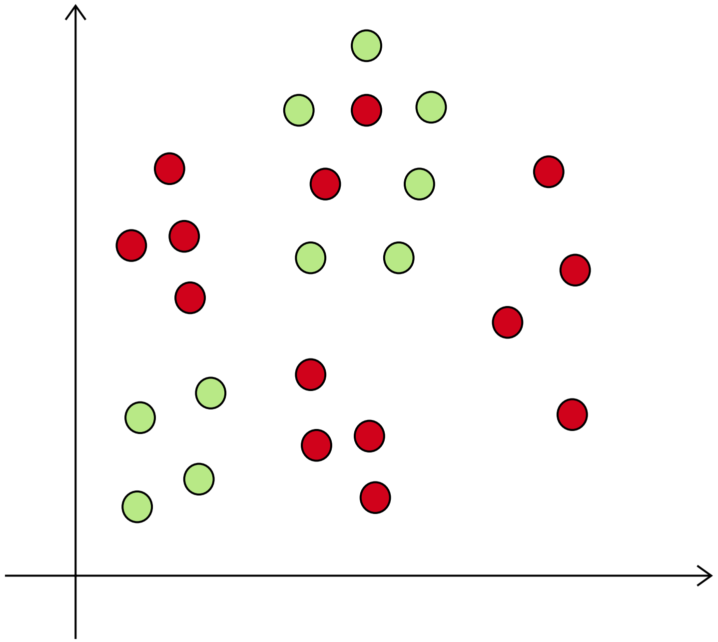
$$\widehat{y} = \begin{cases} 1, & \sigma\left(\mathbf{w}^T\mathbf{x}_i\right) \geqslant T \\ 0, & \sigma\left(\mathbf{w}^T\mathbf{x}_i\right) < T \end{cases} \qquad \widehat{y} = \begin{cases} 1, & \mathbf{w}^T\mathbf{x}_i \geqslant 0 \\ 0, & \mathbf{w}^T\mathbf{x}_i < 0 \end{cases}$$

$$\widehat{\mathbf{y}} = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}$$

Q-17

Choose the model that can be trained to achieve zero training error on all datasets in $\mathbb{R}^2$:

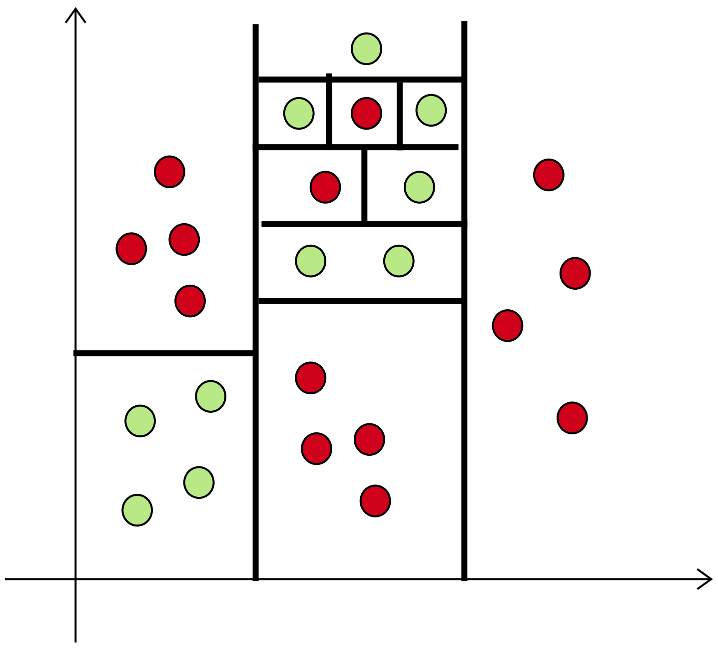
- Logistic regression
- Soft-margin Linear-SVM
- Soft-margin Kernel-SVM with cubic kernel
- Decision Tree



Q-17

Choose the model that can be trained to achieve zero training error on all datasets in $\mathbb{R}^2$:

- Logistic regression
- Soft-margin Linear-SVM
- Soft-margin Kernel-SVM with cubic kernel
- Decision Tree



Q-18

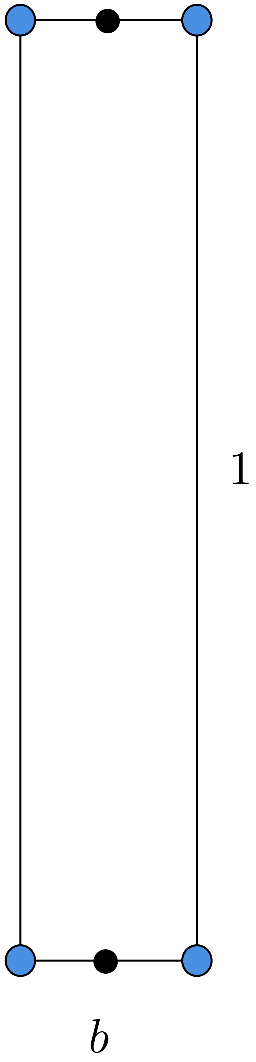
- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration

Q-18

- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration

Q-18

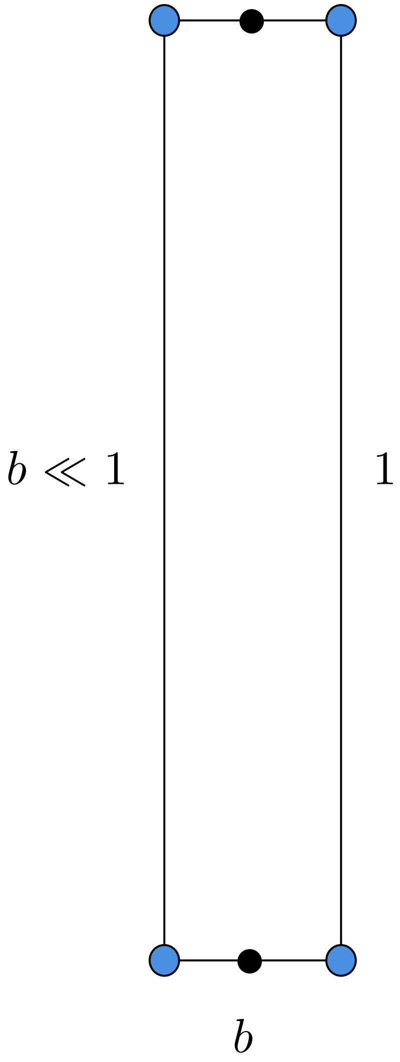
- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration



$$\left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 = b^2$$

Q-18

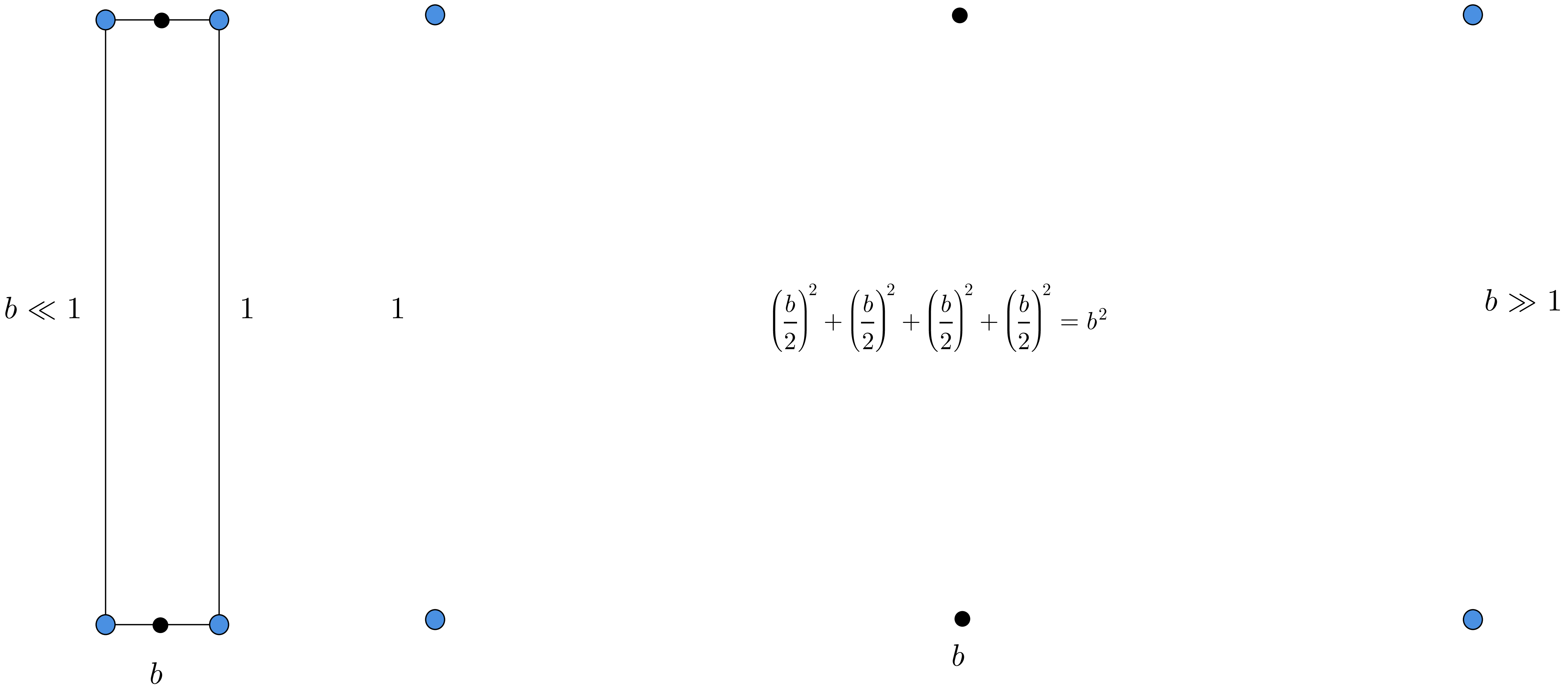
- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration



$$\left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 + \left(\frac{b}{2}\right)^2 = b^2$$

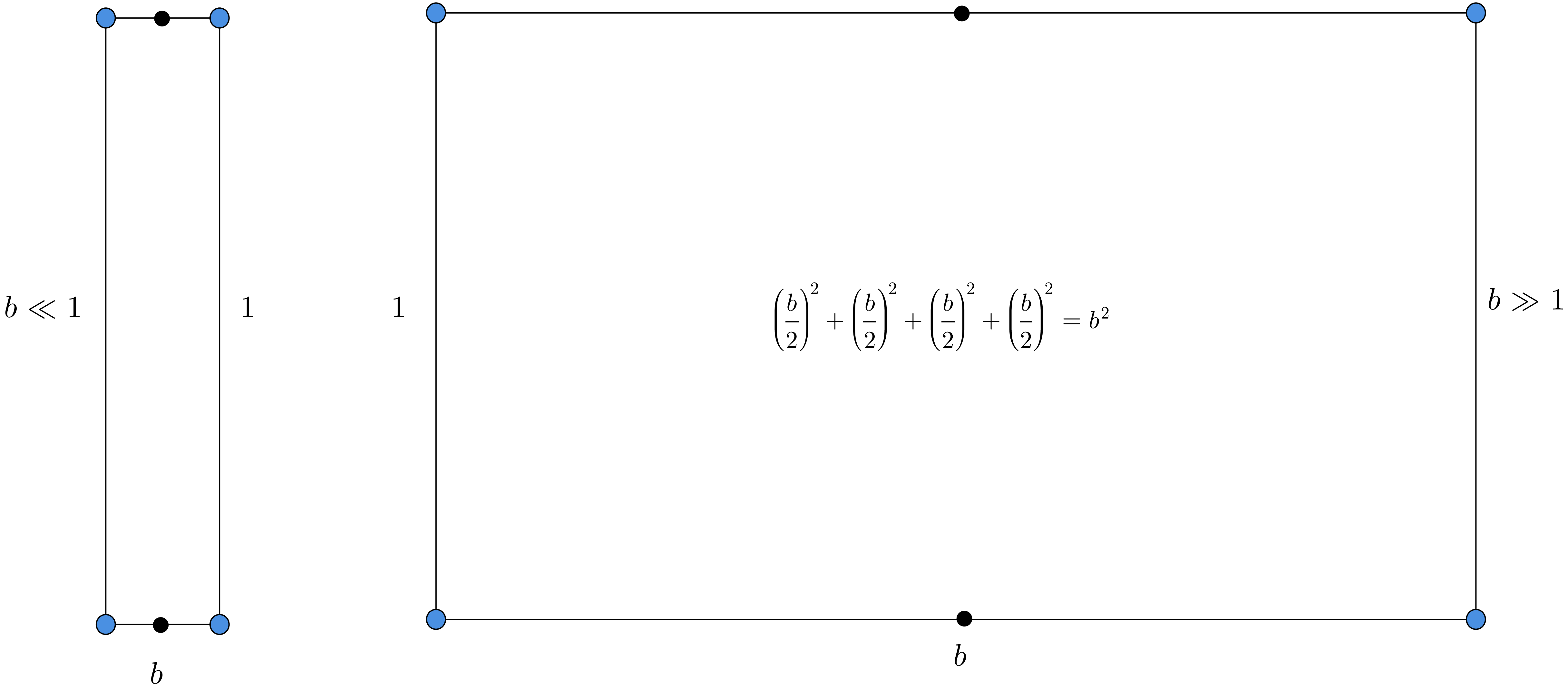
Q-18

- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration



Q-18

- Choose initial centers as mid-point of top two and bottom two data-points
- Run Lloyd's with $k = 2$
- Find the loss for this configuration



Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

$$\propto \text{Beta}(6, 10)$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

$$\log\big(\texttt{posterior}\big) = 5\log p + 9\log(1-p)$$

$$\texttt{Posterior} \propto \texttt{Prior} \times \texttt{Likelihood}$$

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

$$\propto \texttt{Beta}(6, 10)$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

$$\log(\texttt{posterior}) = 5\log p + 9\log(1-p)$$

$$\widehat{p}_{MAP} = \arg\max_p \ \texttt{posterior}(p)$$

$$\texttt{Posterior} \propto \texttt{Prior} \times \texttt{Likelihood}$$

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

$$\propto \texttt{Beta}(6, 10)$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

$$\propto \text{Beta}(6, 10)$$

$$\log\big(\text{posterior}\big) = 5\log p + 9\log(1-p)$$

$$\widehat{p}_{MAP} = \arg\max_p \ \text{posterior}(p)$$

$$\frac{5}{p} - \frac{9}{1-p} = 0$$

$$5 - 5p - 9p = 0$$

$$\widehat{p}_{MAP} = \frac{5}{14}$$

Q-19

- $D = \{1, 0, 1, 0, 1, 0\}$
- Prior is Beta$(3, 7)$
- Find the MAP estimate

Posterior $\propto$ Prior $\times$ Likelihood

$$\propto \frac{p^2(1-p)^6}{B(3,7)} \times p^3(1-p)^3$$

$$\propto p^5(1-p)^9$$

$$\propto \text{Beta}(6, 10)$$

$$\log\big(\text{posterior}\big) = 5\log p + 9\log(1-p)$$

$$\widehat{p}_{MAP} = \arg\max_p \ \text{posterior}(p)$$

$$\frac{5}{p} - \frac{9}{1-p} = 0$$

$$5 - 5p - 9p = 0$$

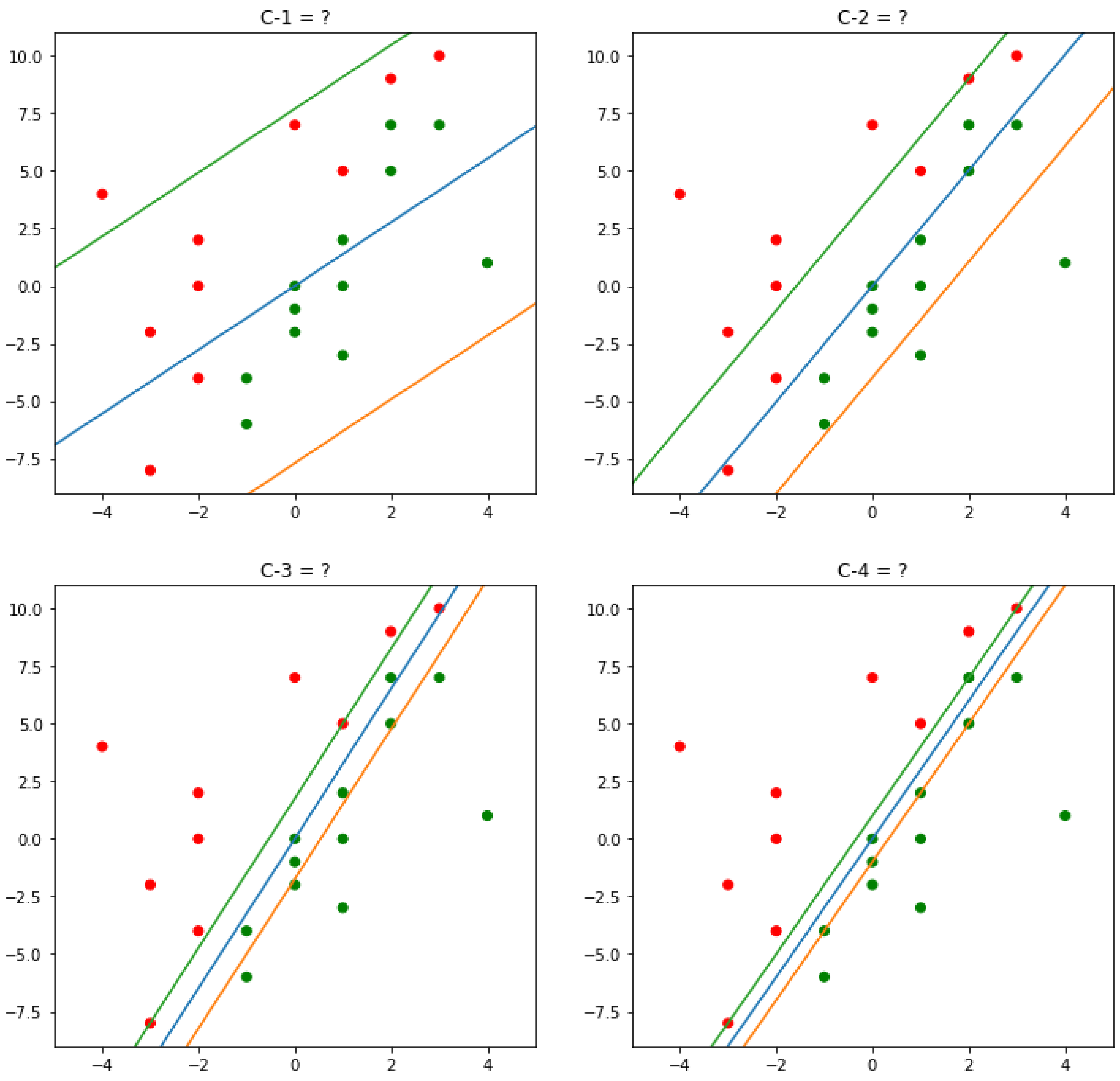
$$\widehat{p}_{MAP} = \frac{5}{14}$$

If $\alpha, \beta > 1$, then mode of Beta(α, β) is:

$$\frac{\alpha - 1}{\alpha + \beta - 2}$$

Q-20

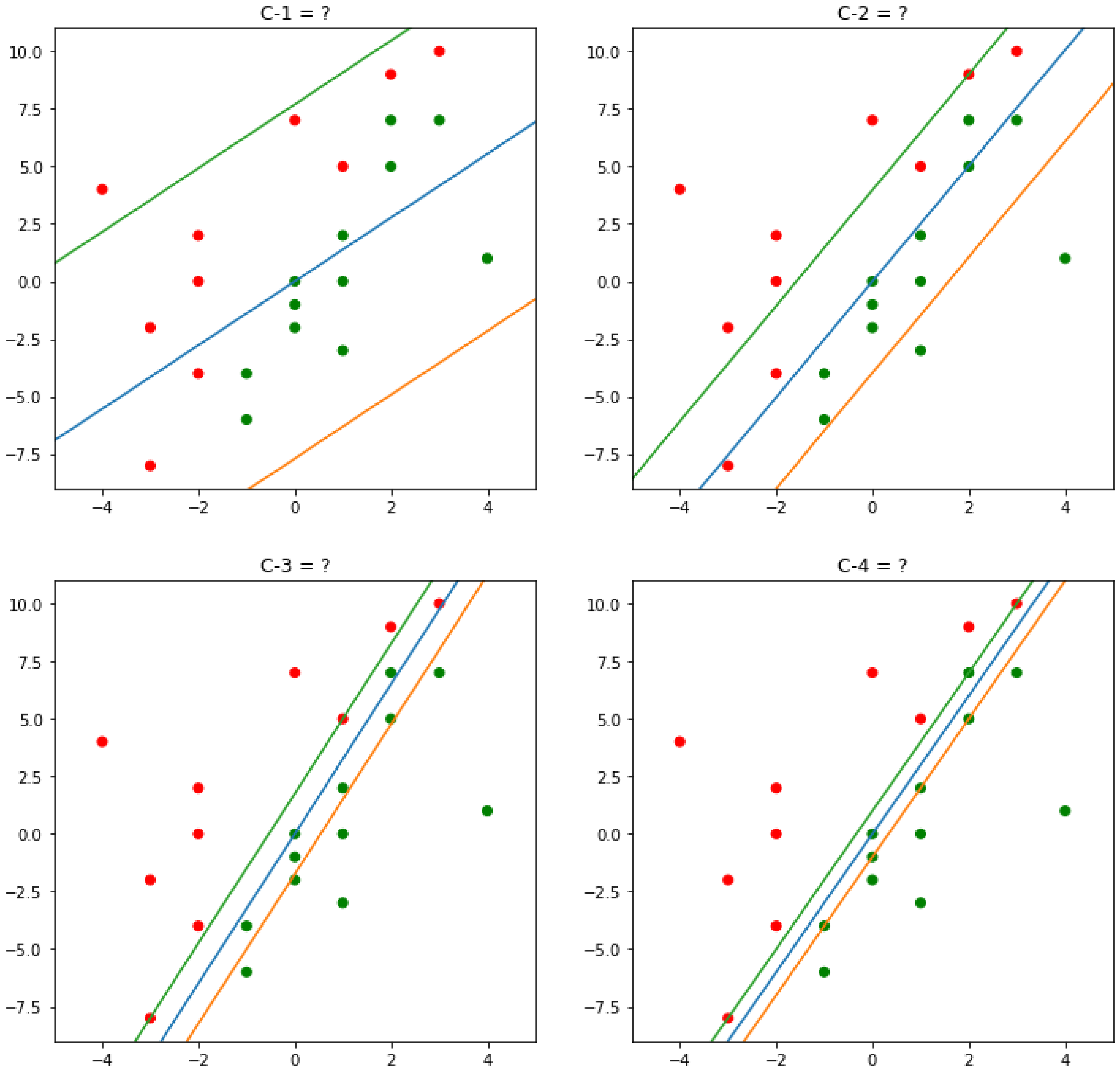
- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$



Q-20

- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$

$$\min_{\mathbf{w}} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n} \max\left[0, 1-\left(\mathbf{w}^T\mathbf{x}_i\right)y_i\right]$$

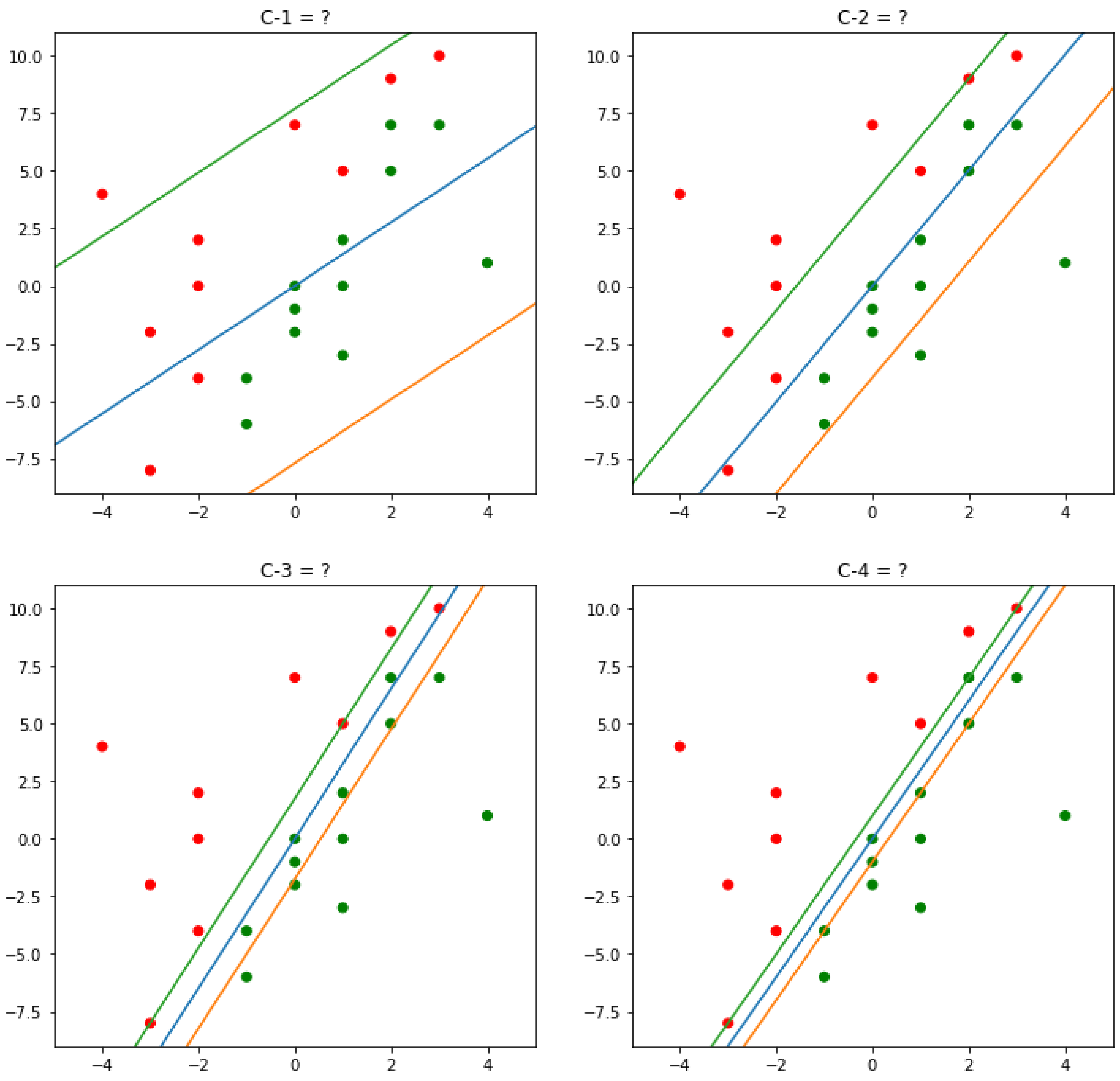


Q-20

- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$

$$\min_{\mathbf{w}} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n} \max\left[0, 1-\left(\mathbf{w}^T\mathbf{x}_i\right)y_i\right]$$

Loss = Margin + Hinge-loss



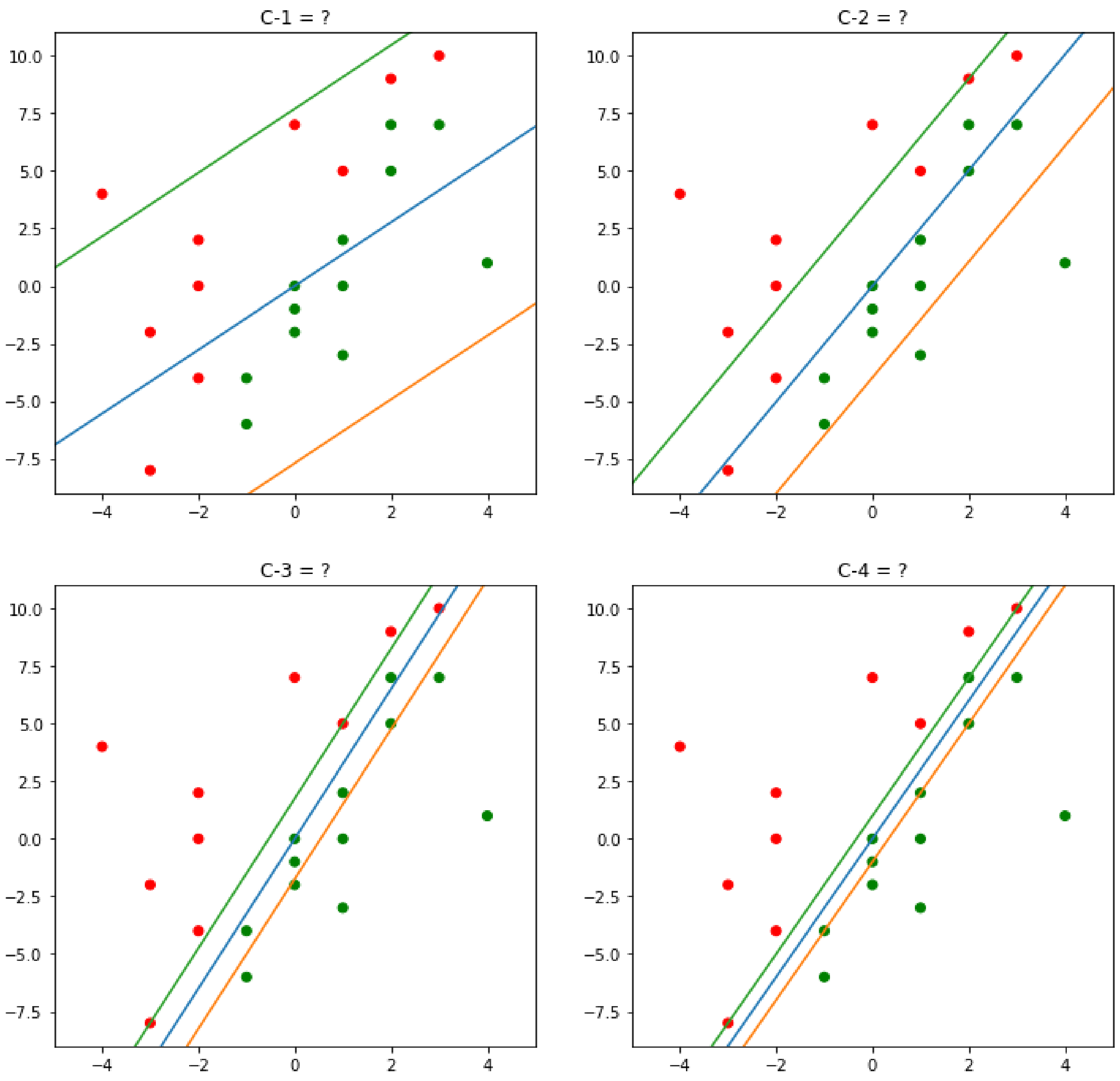
Q-20

- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$

$$\min_{\mathbf{w}} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n} \max\left[0, 1 - \left(\mathbf{w}^T\mathbf{x}_i\right)y_i\right]$$

Loss = Margin + Hinge-loss

Ideal: Small $||\mathbf{w}||$ and small hinge loss



Q-20

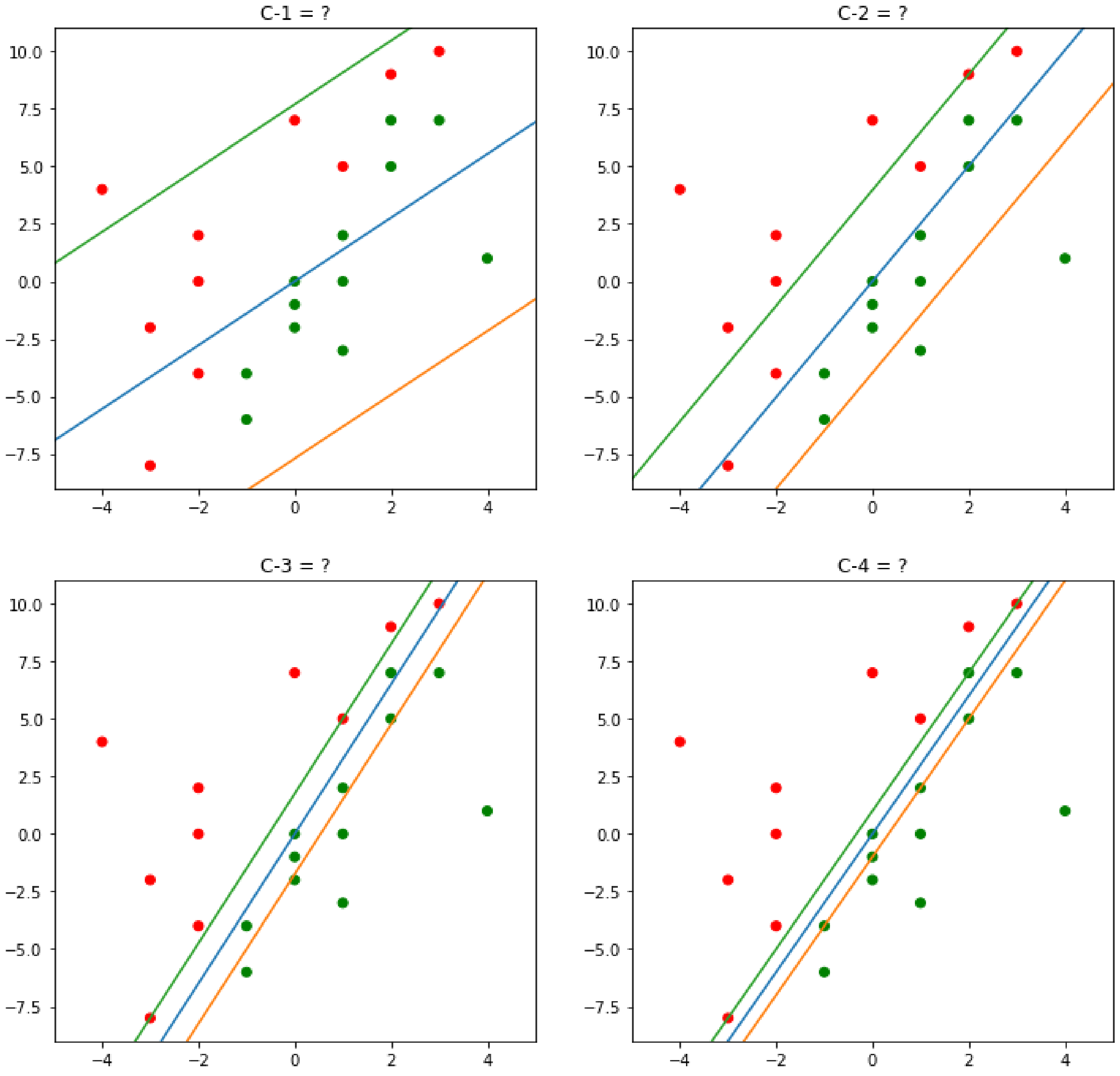
- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$

$$\min_{\mathbf{w}} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n} \max\left[0, 1 - \left(\mathbf{w}^T\mathbf{x}_i\right)y_i\right]$$

Loss = Margin + Hinge-loss

Ideal: Small $||\mathbf{w}||$ and small hinge loss

Large $||\mathbf{w}|| \implies$ Narrow margin $\implies$ Small hinge loss



Q-20

- Soft-Margin Linear-SVM
- Map C_i to $[0.01,\ 0.1,\ 1,\ 10]$ for $1 \leqslant i \leqslant 4$

$$\min_{\mathbf{w}} \quad \frac{||\mathbf{w}||^2}{2} + C\sum_{i=1}^{n} \max\left[0, 1 - \left(\mathbf{w}^T\mathbf{x}_i\right)y_i\right]$$

Loss = Margin + Hinge-loss

Ideal: Small $||\mathbf{w}||$ and small hinge loss

Large $||\mathbf{w}||$ $\Longrightarrow$ Narrow margin $\Longrightarrow$ Small hinge loss

Small $||\mathbf{w}||$ $\Longrightarrow$ Wide margin $\Longrightarrow$ Large hinge loss