CSC 276: Data Science Lecture #06

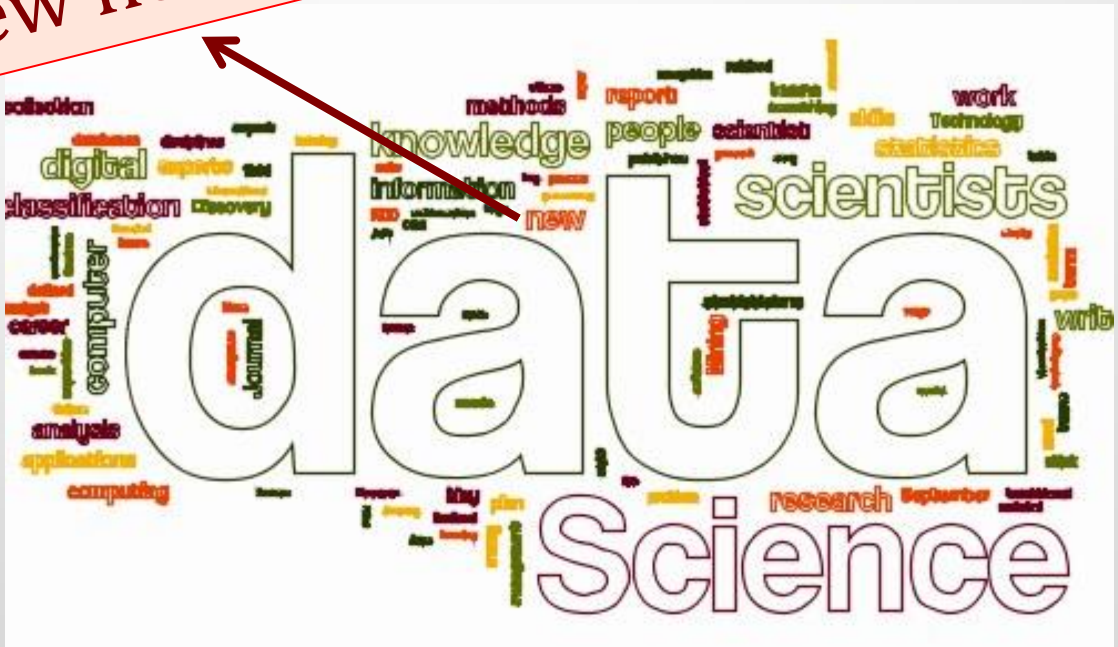
Dr.Fatema Nafa

Fall 2022

12/30/2022

Data Science

This class is truly seminar-style: I'm here, as you are, in order to gain insights into this very new field… .



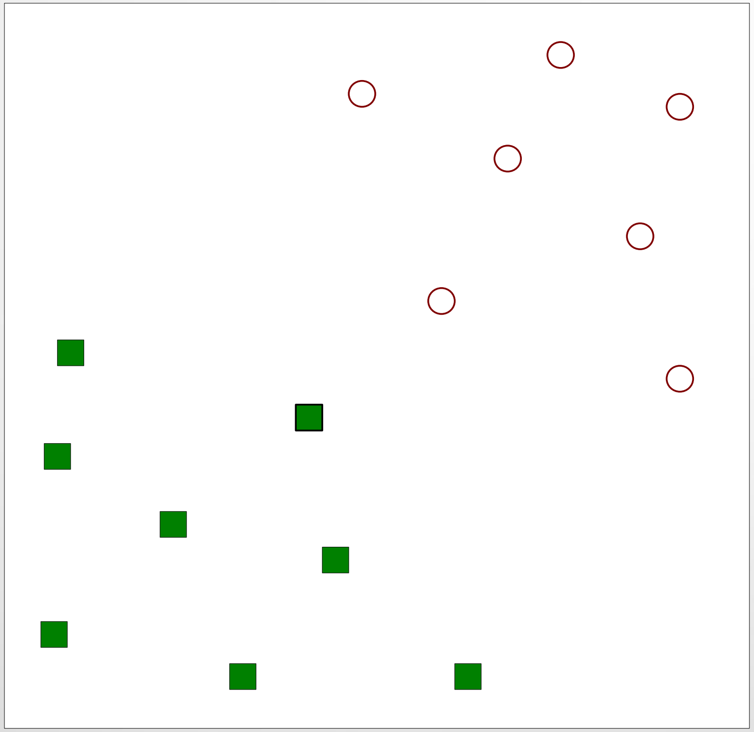
Course Information

Instructor	Dr. Fatema Nafa
Email	fnafa@Salemstate.edu
Office Hours	Tuesdays &Wednesdays 2:00 PM – 3:00 PM & by appointment
Lecture Time	Tuesdays & Thursdays: 12:15 PM – 01:30 PM
Place	Meier 346
Final Exam: Tuesday December 21 11:00 – 1:00	


Any
Questions

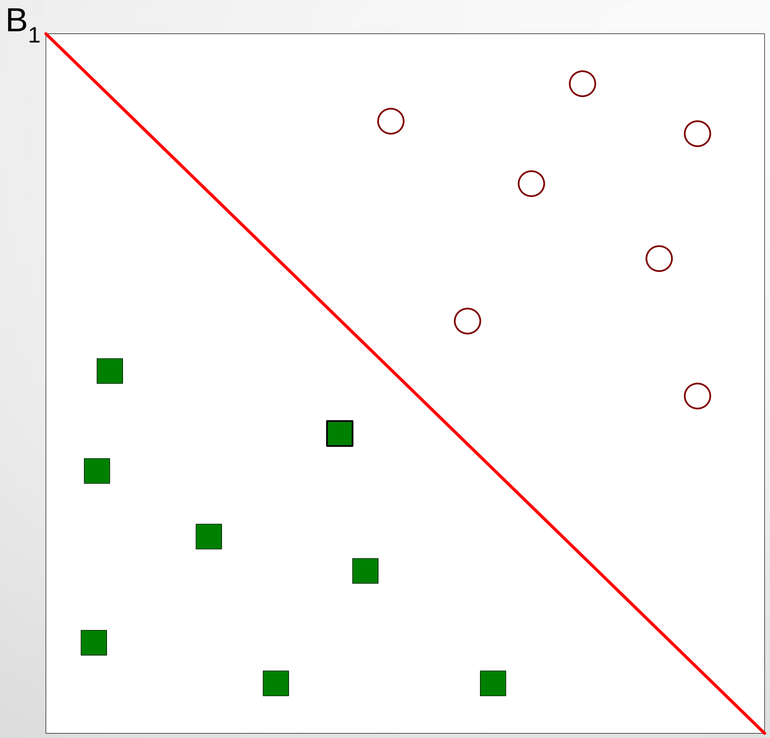
Support Vector Machines

Support Vector Machines



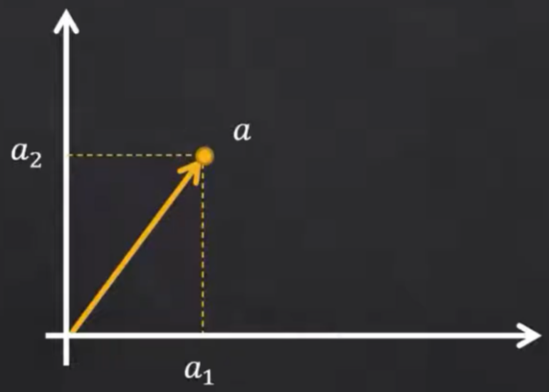
- Find a linear hyperplane (decision boundary) that will separate the data

Support Vector Machines

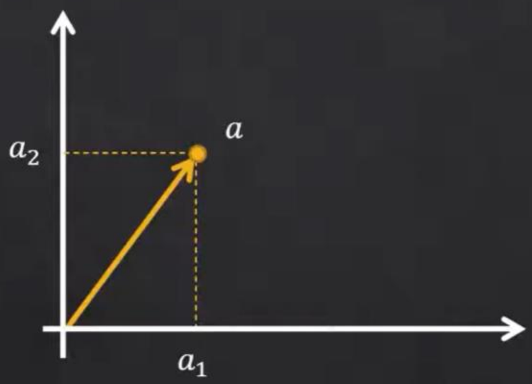


- One Possible Solution

Properties of vectors



Properties of vectors

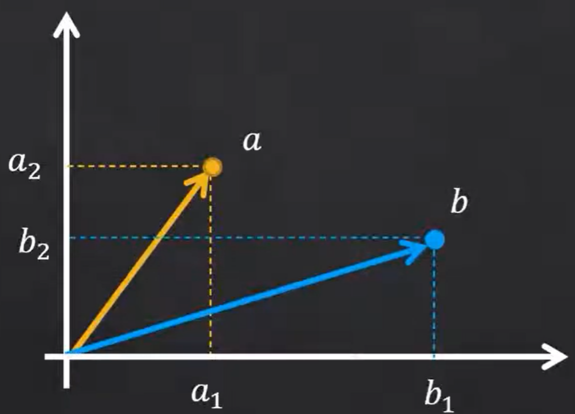


$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

Norm: $\|a\| = \sqrt{a_1^{\ 2} + a_2^{\ 2}}$

Direction: $\dfrac{\boldsymbol{a}}{\|a\|}$

Properties of vectors

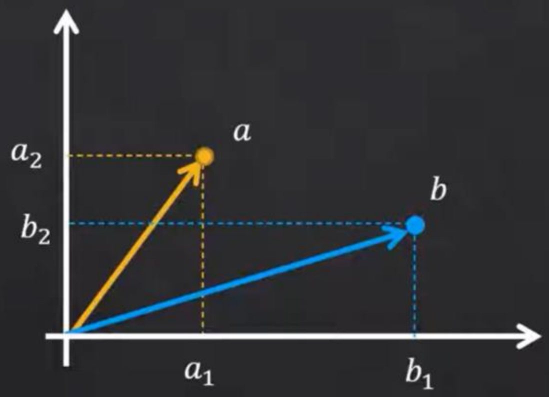


$$\boldsymbol{a}=\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix} \ \\ \ \end{bmatrix}$$

Norm: $\|a\| = \sqrt{{a_1}^2 + {a_2}^2}$

Direction: $\dfrac{\boldsymbol{a}}{\|a\|}$

Properties of vectors



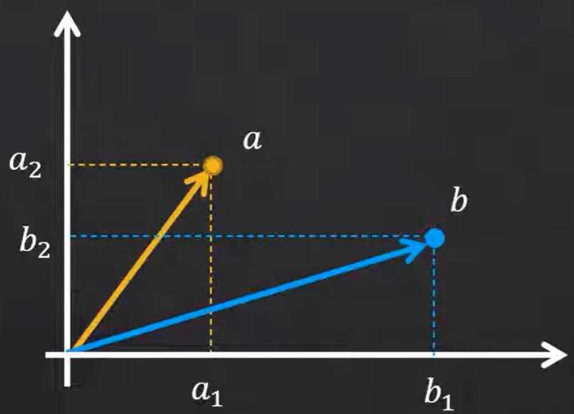
$$\boldsymbol{a}=\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Norm: $\|a\| = \sqrt{{a_1}^2 + {a_2}^2}$

Direction: $\frac{\boldsymbol{a}}{\|a\|}$

Addition: $\boldsymbol{a}+\boldsymbol{b} = (a_1 + b_1,\ a_2 + b_2)$

Properties of vectors



$$\boldsymbol{a}=\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

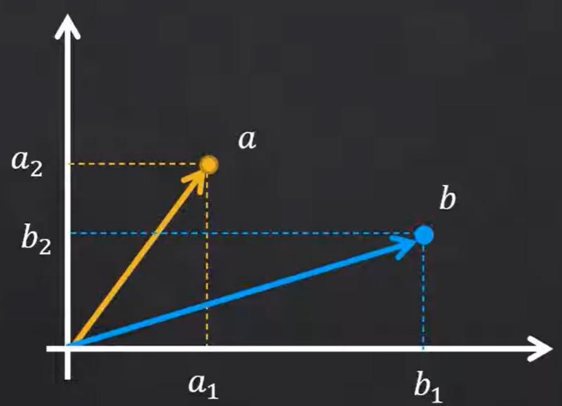
Norm: $\|a\| = \sqrt{a_1{}^2 + a_2{}^2}$

Direction: $\frac{\boldsymbol{a}}{\|a\|}$

Addition: $\boldsymbol{a}+\boldsymbol{b} = (a_1 + b_1,\ a_2 + b_2)$

Subtraction: $\boldsymbol{a}\text{-}\boldsymbol{b} = (a_1 - b_1,\ a_2 - b_2)$

Properties of vectors



$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Norm: $\|a\| = \sqrt{a_1{}^2 + a_2{}^2}$

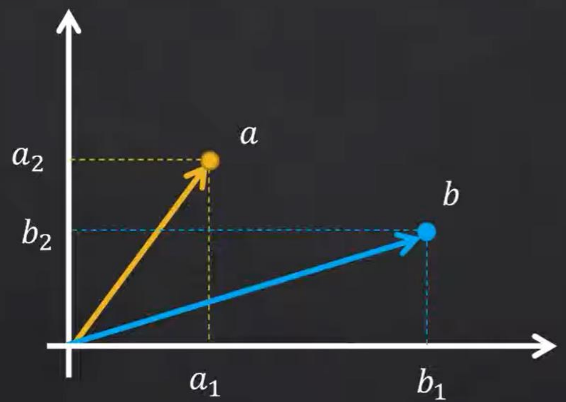
Direction: $\dfrac{\boldsymbol{a}}{\|a\|}$

Addition: $\boldsymbol{a}+\boldsymbol{b} = (a_1 + b_1,\ a_2 + b_2)$

Subtraction: $\boldsymbol{a}\text{-}\boldsymbol{b} = (a_1 - b_1,\ a_2 - b_2)$

$$\boldsymbol{a}^T\boldsymbol{b} = ?$$

Properties of vectors



$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Norm: $\|a\| = \sqrt{{a_1}^2 + {a_2}^2}$

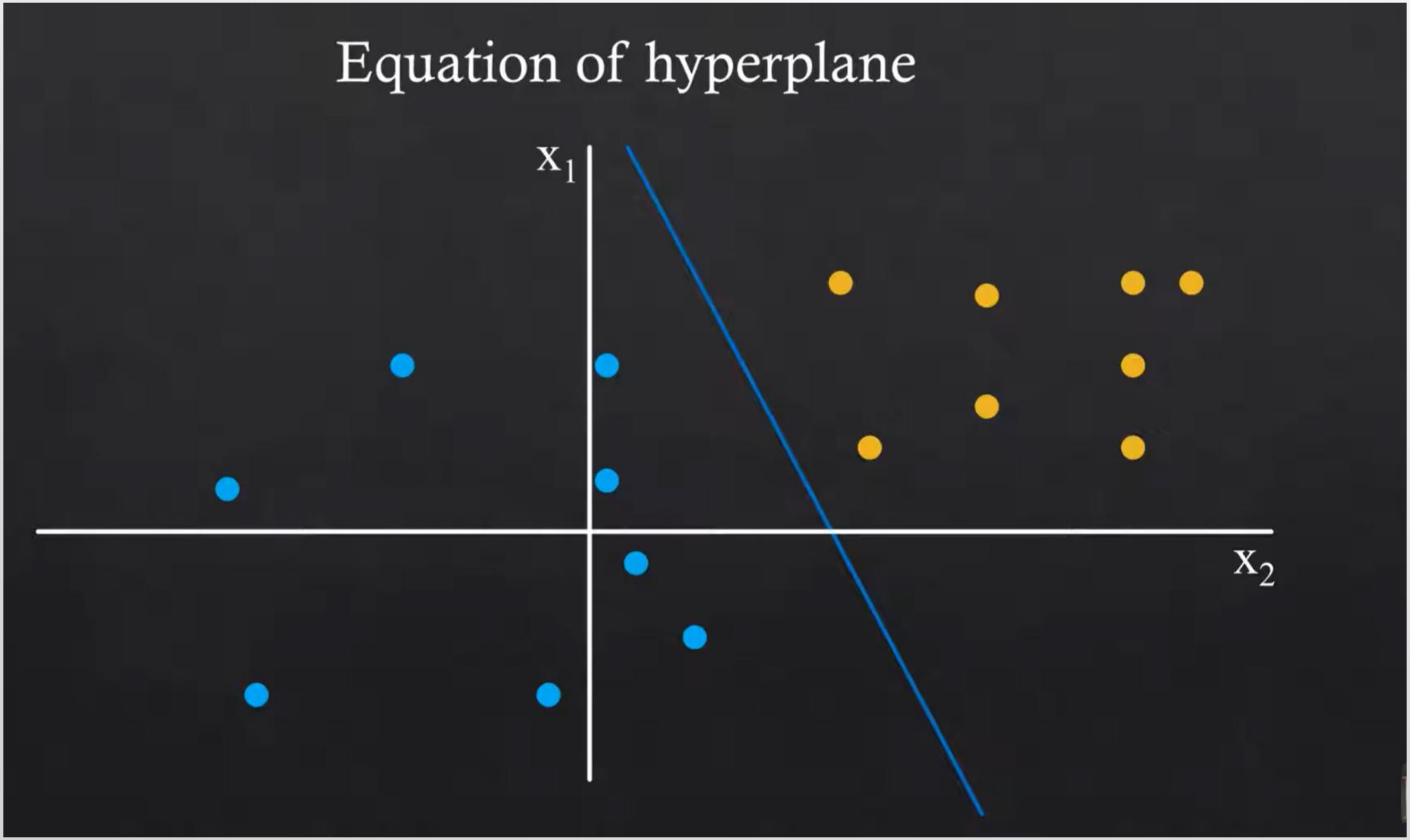
Direction: $\dfrac{\boldsymbol{a}}{\|a\|}$

Addition: $\boldsymbol{a}+\boldsymbol{b} = (a_1 + b_1,\ a_2 + b_2)$

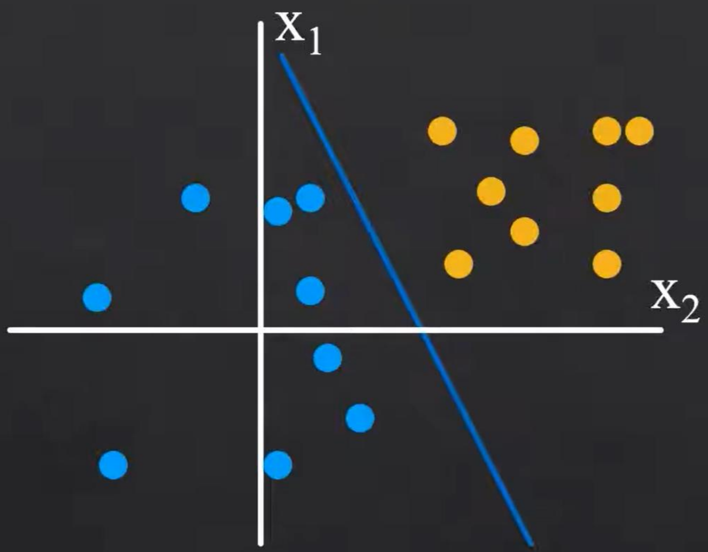
Subtraction: $\boldsymbol{a}\text{-}\boldsymbol{b} = (a_1 - b_1,\ a_2 - b_2)$

$$\boldsymbol{a}^T\boldsymbol{b} = ?$$

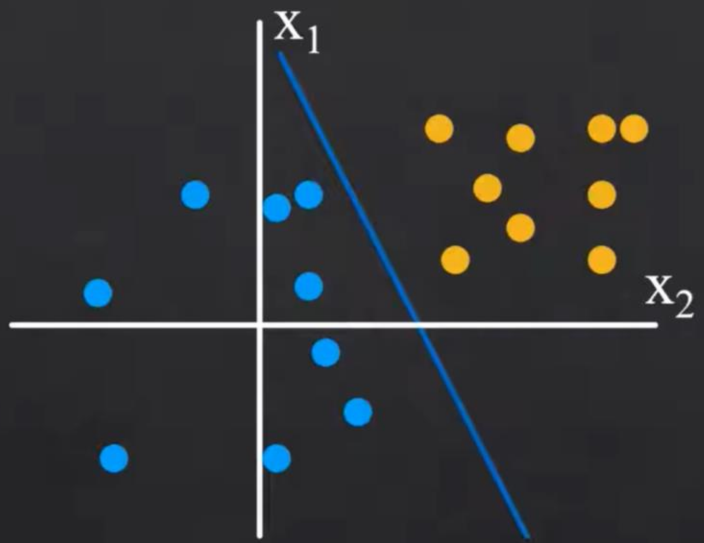
inner product: $\boldsymbol{a}^T\boldsymbol{b} = a_1 \times b_1 + a_2 \times b_2$
$= \boldsymbol{b}^T\boldsymbol{a}$



Equation of hyperplane
x1
x2

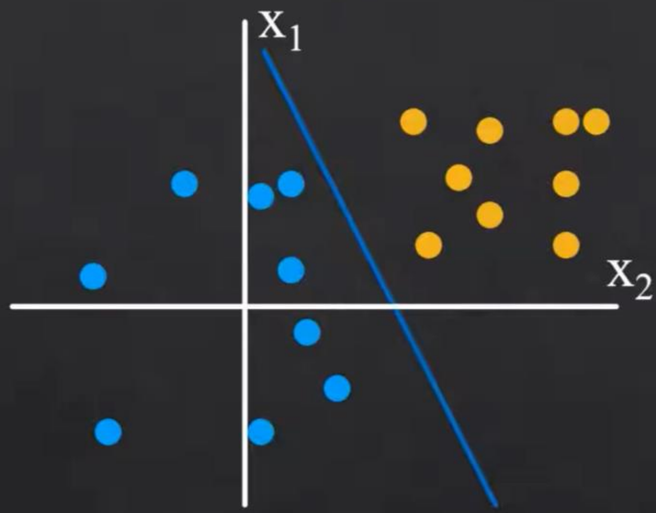


Line Equation : $y = m \times x + b$



Line Equation :

$$y = m \times x + b$$
$$m \times x - y + b = 0$$

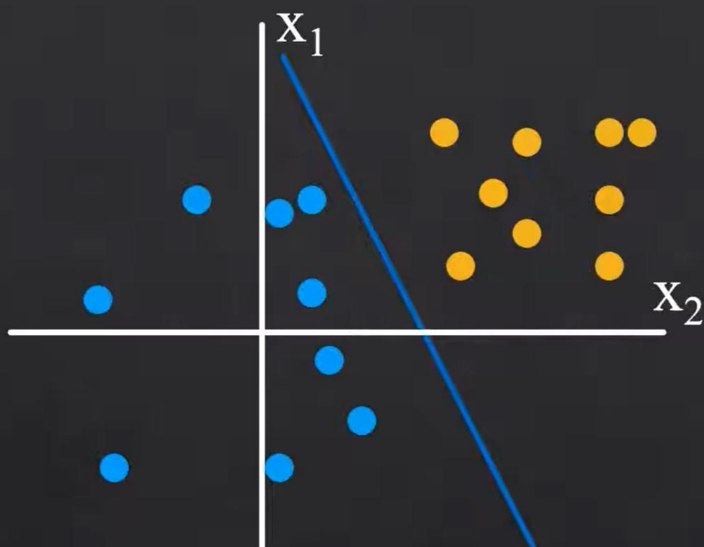


Line Equation :

$$y = m \times x + b$$
$$m \times x - y + b = 0$$

We define two vectors,

$\boldsymbol{w} = \begin{pmatrix} m \\ -1 \end{pmatrix}$ and $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$



Line Equation :

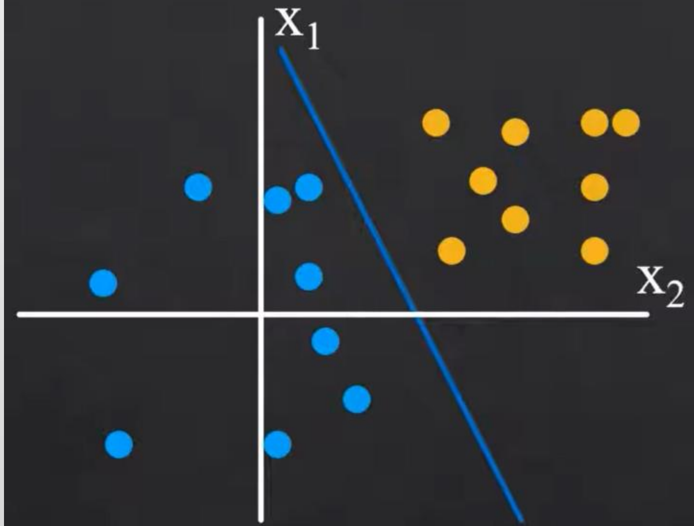
$$y = m \times x + b$$

$$m \times x - y + b = 0$$

We define two vectors,

$$w = \begin{pmatrix} m \\ -1 \end{pmatrix} \text{ and } x = \begin{pmatrix} x \\ y \end{pmatrix}$$

$$w^T.x = m \times x + (-1) \times y$$



Line Equation :

$$y = m \times x + b$$
$$m \times x - y + b = 0$$

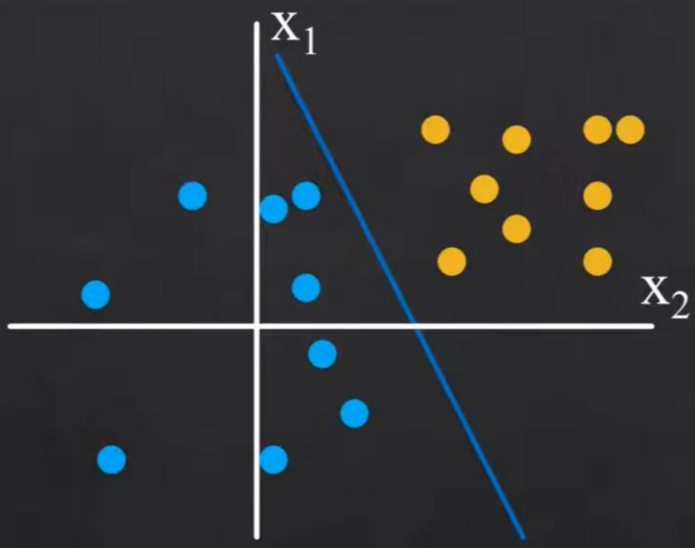
We define two vectors,

$\boldsymbol{w} = \begin{pmatrix} m \\ -1 \end{pmatrix}$ and $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$

$$\boldsymbol{w}^T.\boldsymbol{x} = m \times x + (-1) \times y$$

So that,

$$\boldsymbol{w}^T.\boldsymbol{x} + b = 0$$



Line Equation : $y = m \times x + b$

$m \times x - y + b = 0$

We define two vectors,

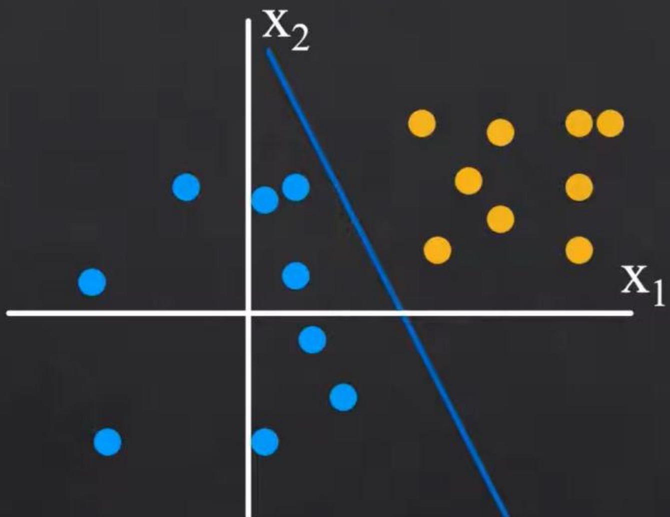
$\boldsymbol{w} = \begin{pmatrix} m \\ -1 \end{pmatrix}$ and $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$

$\boldsymbol{w}^T.\boldsymbol{x} = m \times x + (-1) \times y$

So that,

$\boldsymbol{w}^T.\boldsymbol{x} + b = 0$

Equation of Hyperplane : $\boldsymbol{w}^T.\boldsymbol{x} + b = 0$



Line Equation : $y = m \times x + b$

$m \times x - y + b = 0$

We define two vectors,

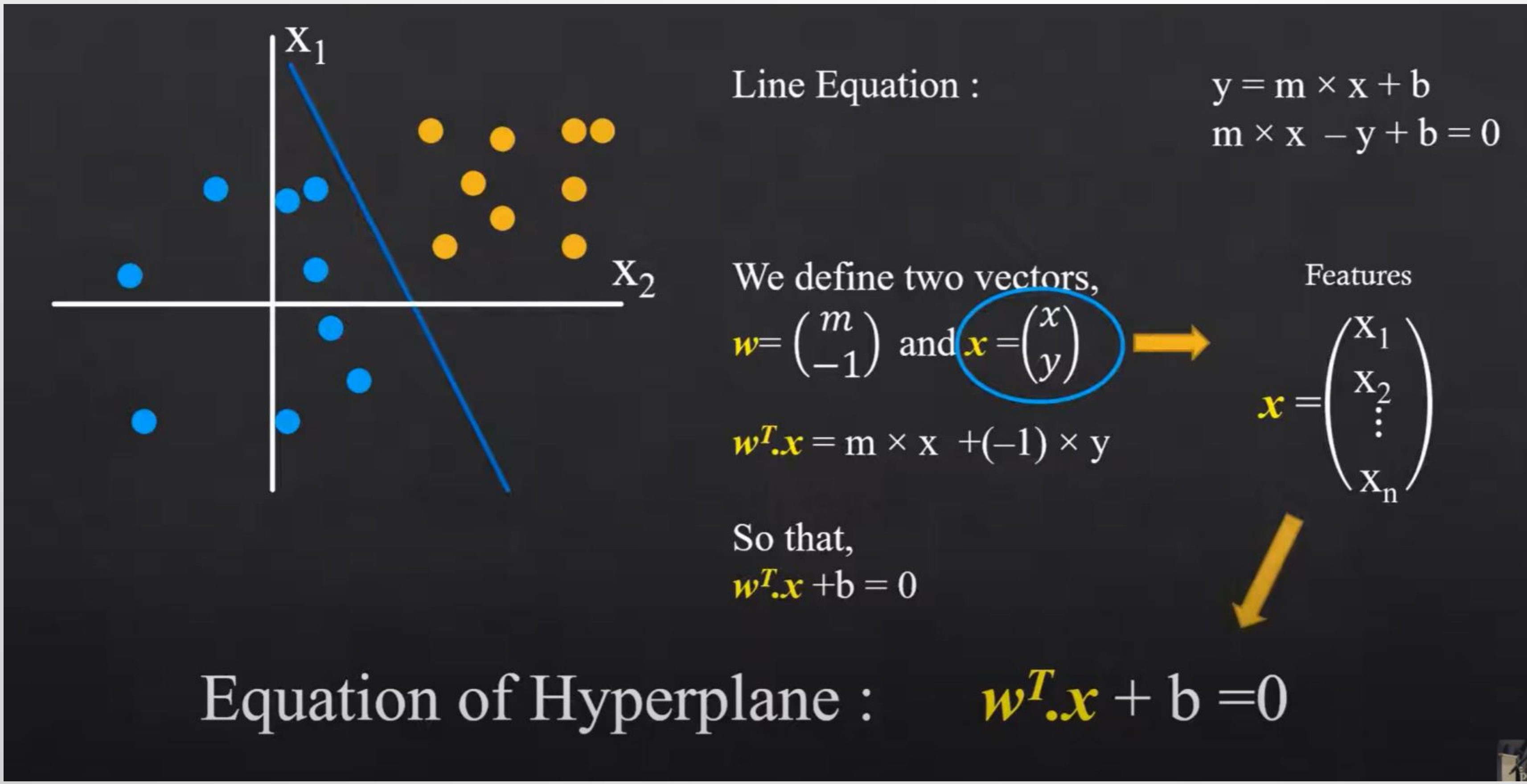
$w = \begin{pmatrix} m \\ -1 \end{pmatrix}$ and $x = \begin{pmatrix} x \\ y \end{pmatrix}$ $\Rightarrow$ Features $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

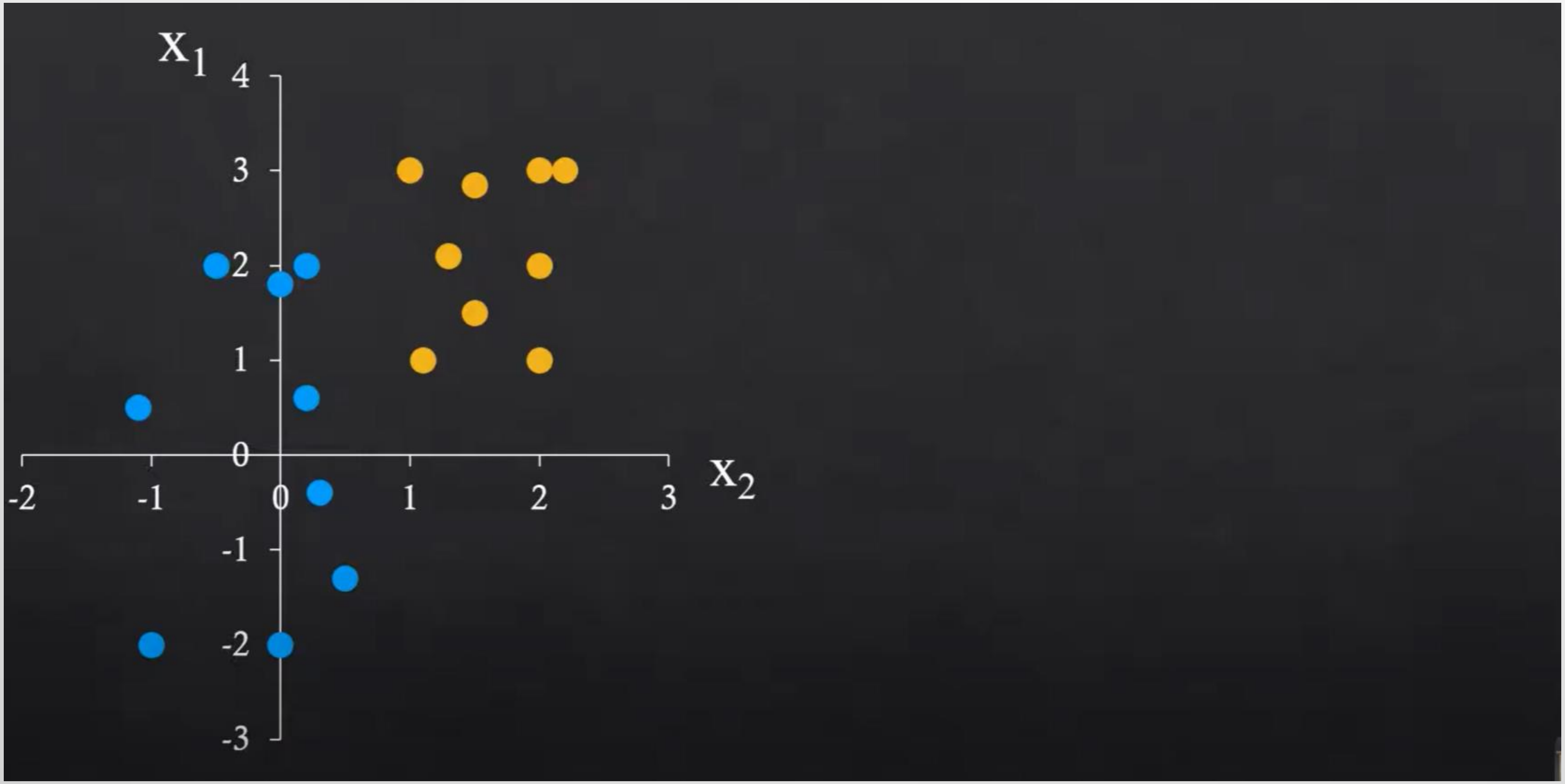
$w^T.x = m \times x + (-1) \times y$

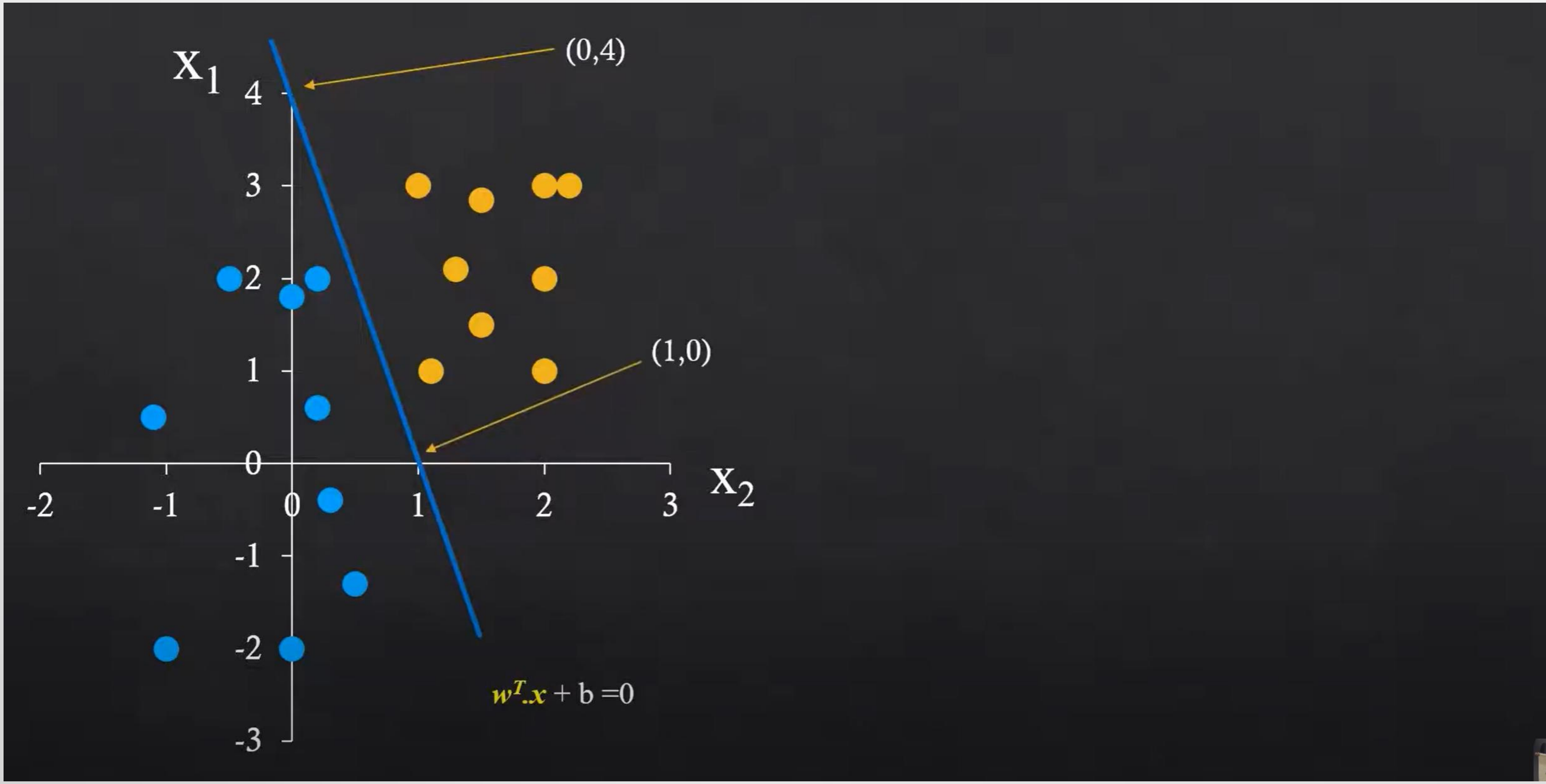
So that,

$w^T.x + b = 0$

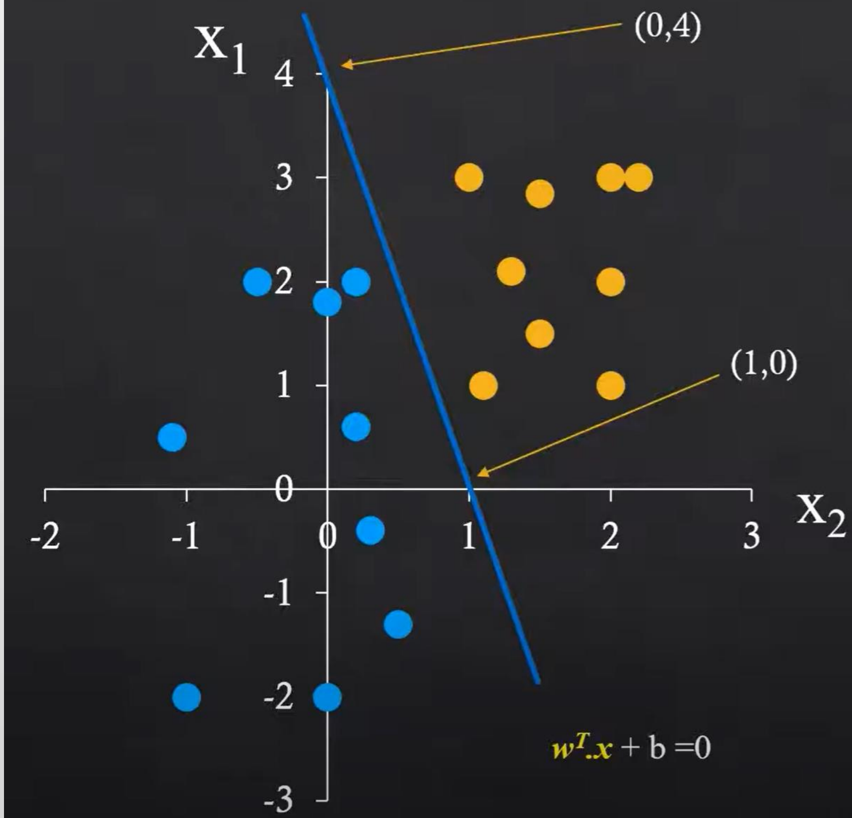
Equation of Hyperplane : $w^T.x + b = 0$


x_1
x_2
Line Equation :
$y = m \times x + b$
$m \times x - y + b = 0$
We define two vectors,
$w = \begin{pmatrix} m \\ -1 \end{pmatrix}$ and $x = \begin{pmatrix} x \\ y \end{pmatrix}$
Features
$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$
$w^T.x = m \times x + (-1) \times y$
So that,
$w^T.x + b = 0$
Equation of Hyperplane : $w^T.x + b = 0$


X1
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
X2

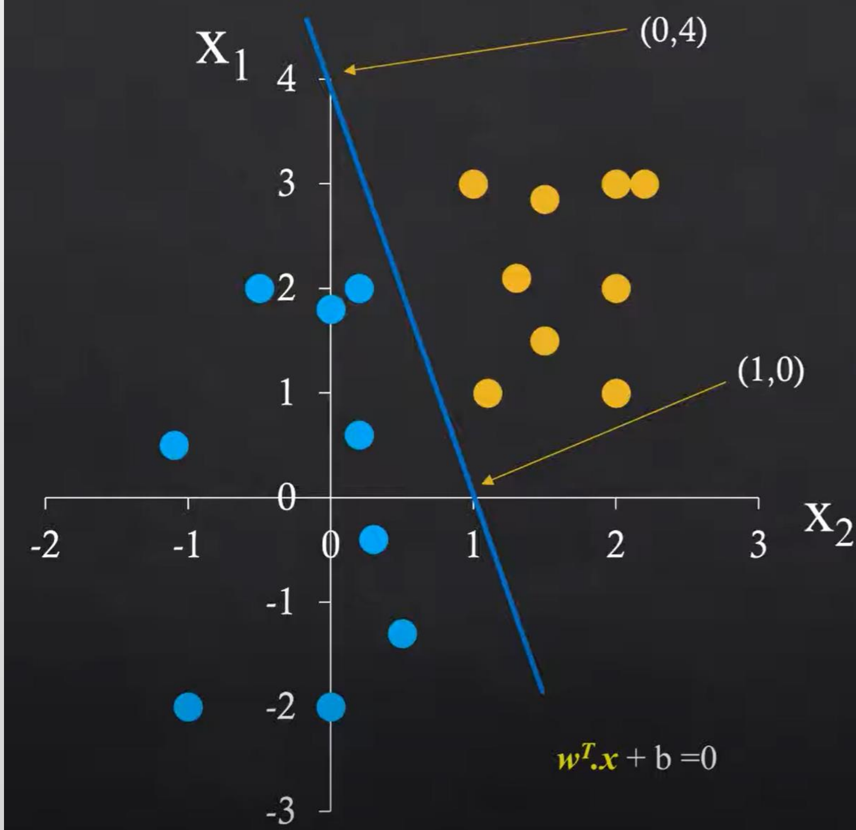


x_1
x_2
(0,4)
(1,0)
4
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
$w^T.x + b = 0$



Slope, m = (4-0)/(0-1) = -4

$$w = \begin{pmatrix} m \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$$



Slope, $m = (4-0)/(0-1) = -4$

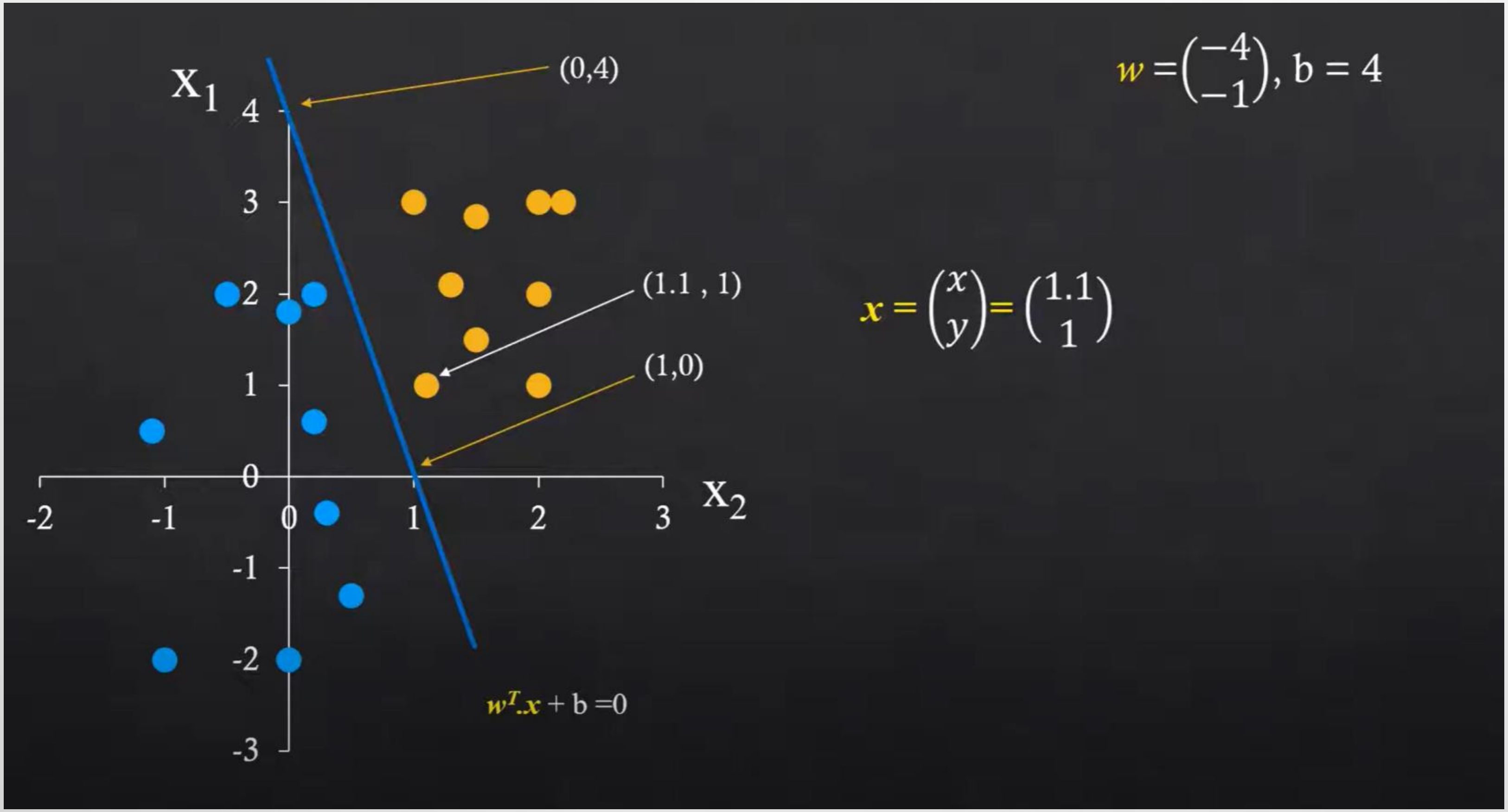
$$w = \begin{pmatrix} m \\ -1 \end{pmatrix} = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$$

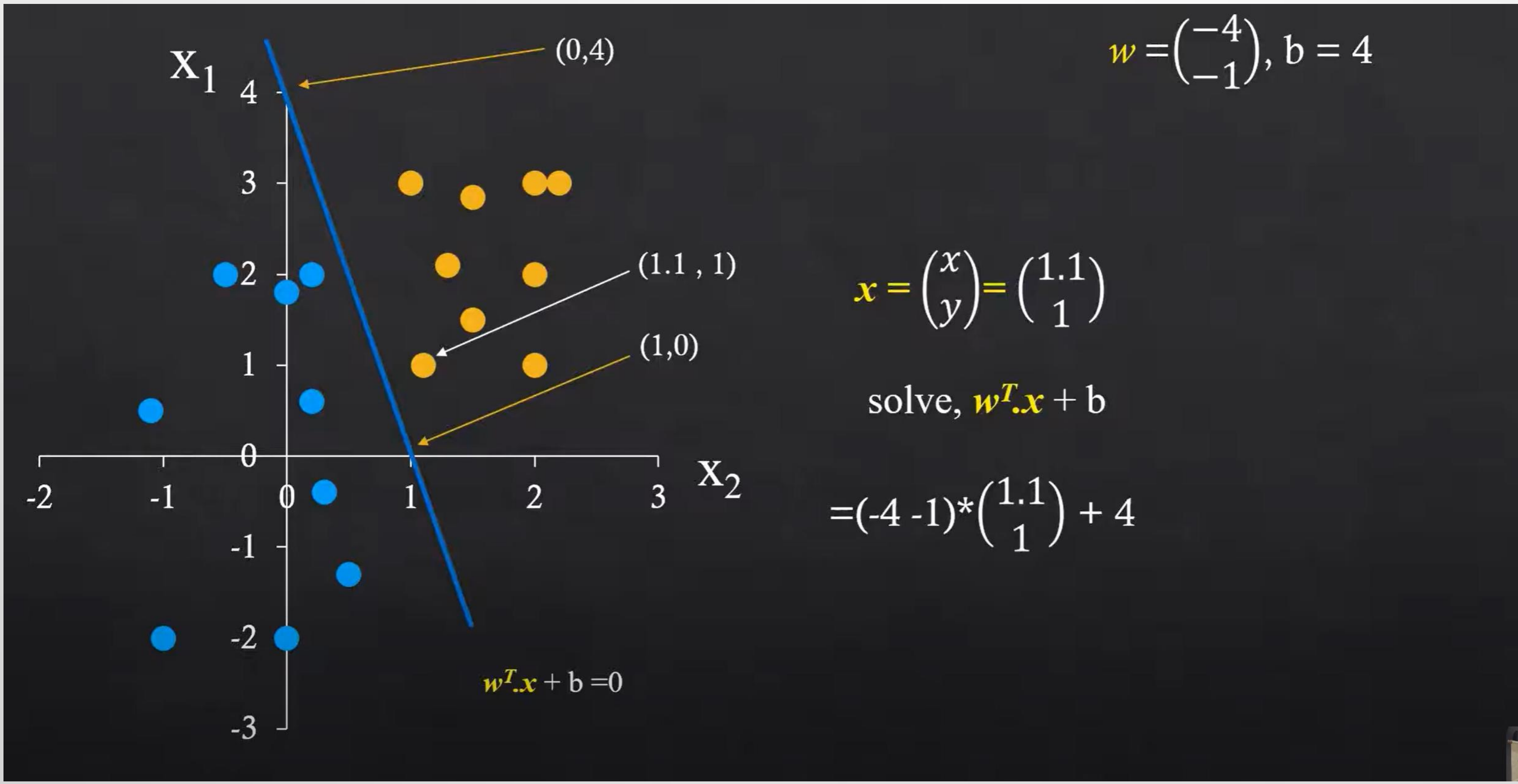
$$x = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

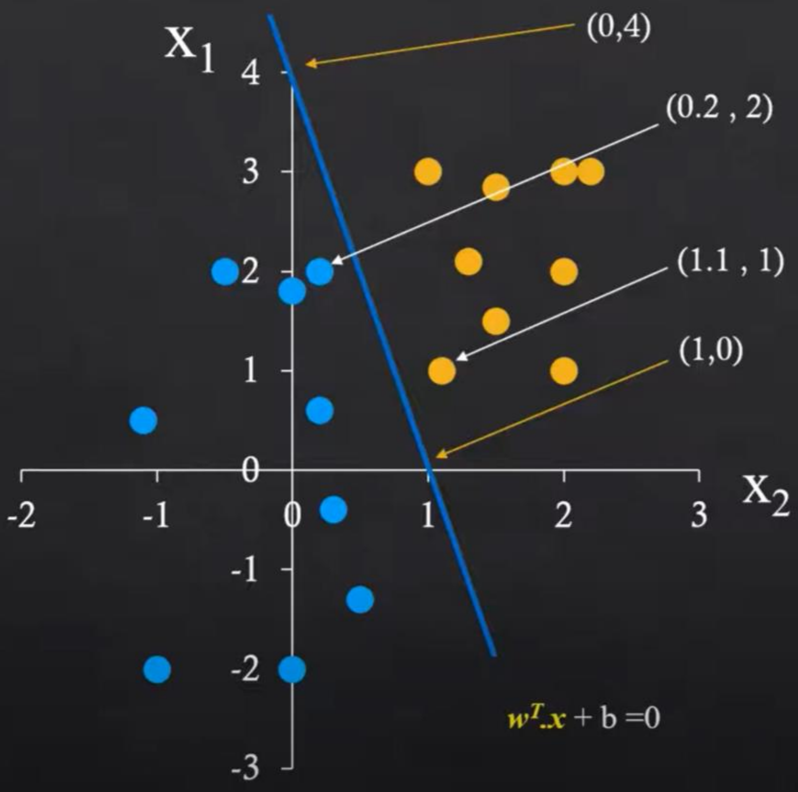
$$w^T.x + b = 0$$

$$(-4\ -1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b = 0$$

$$b = 4$$


$w = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$, b = 4
$x = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.1 \\ 1 \end{pmatrix}$
(0,4)
(1.1 , 1)
(1,0)
x_1
x_2
$w^T.x + b = 0$


X_1
X_2
(0,4)
(1.1 , 1)
(1,0)
$w^T.x + b = 0$
$w = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$, b = 4
$x = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.1 \\ 1 \end{pmatrix}$
solve, $w^T.x + b$
$=(-4\ -1)*\begin{pmatrix} 1.1 \\ 1 \end{pmatrix} + 4$



$$w = \begin{pmatrix}-4\\-1\end{pmatrix},\ b = 4$$

$$x = \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1.1\\1\end{pmatrix}$$

solve, $w^T.x + b$

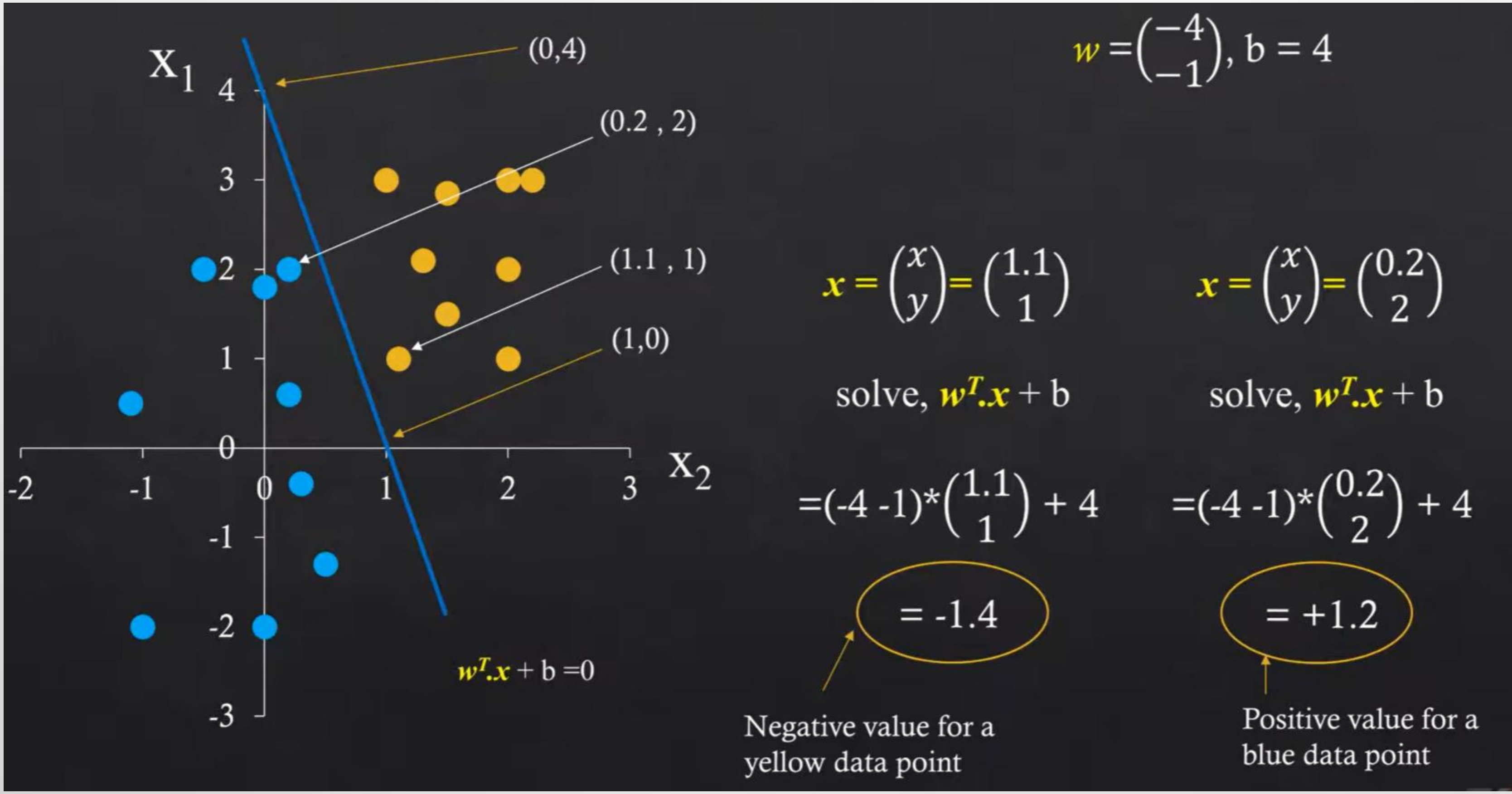
$$=(-4\ -1)*\begin{pmatrix}1.1\\1\end{pmatrix} + 4$$

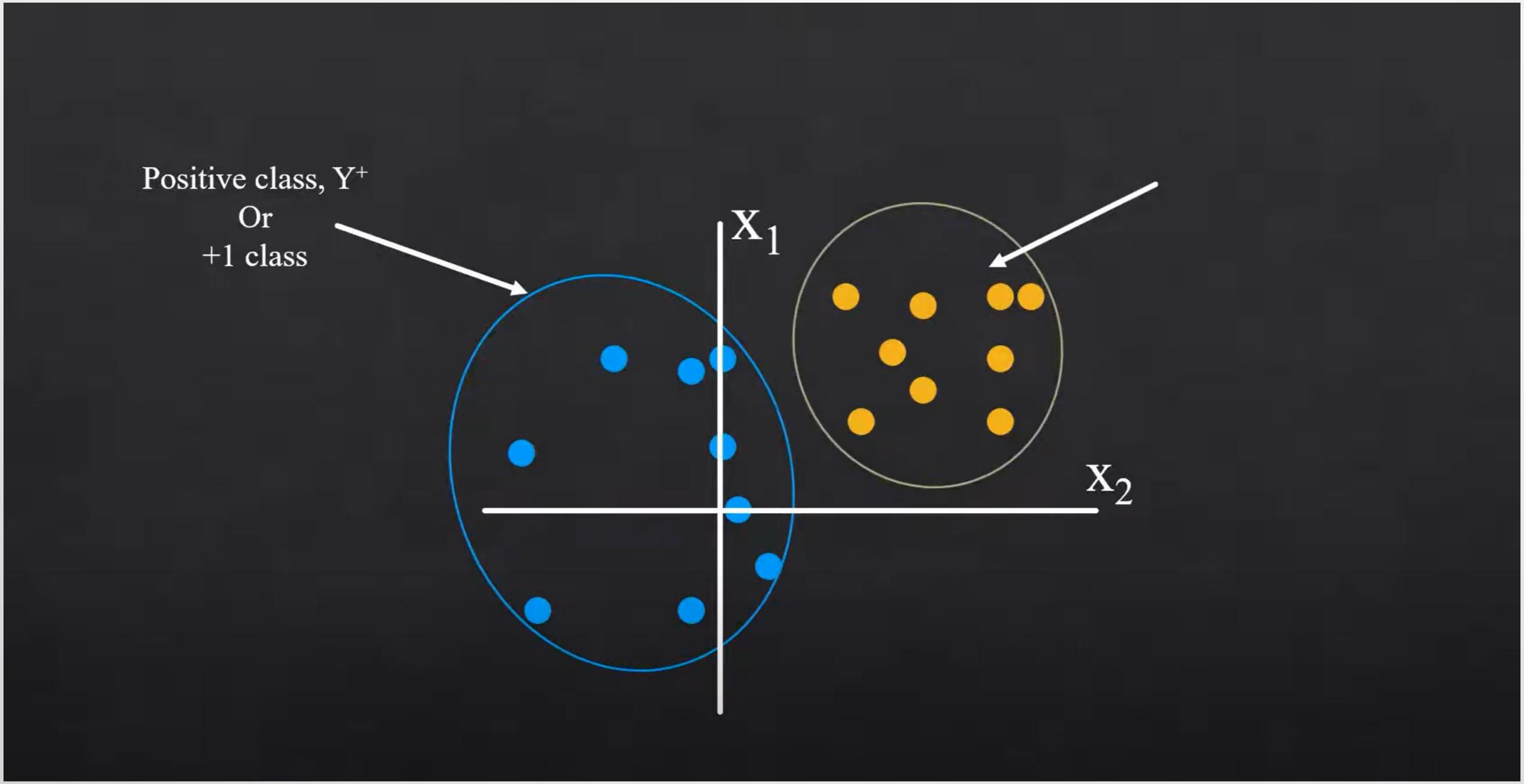
$$= -1.4$$

$$x = \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}0.2\\2\end{pmatrix}$$

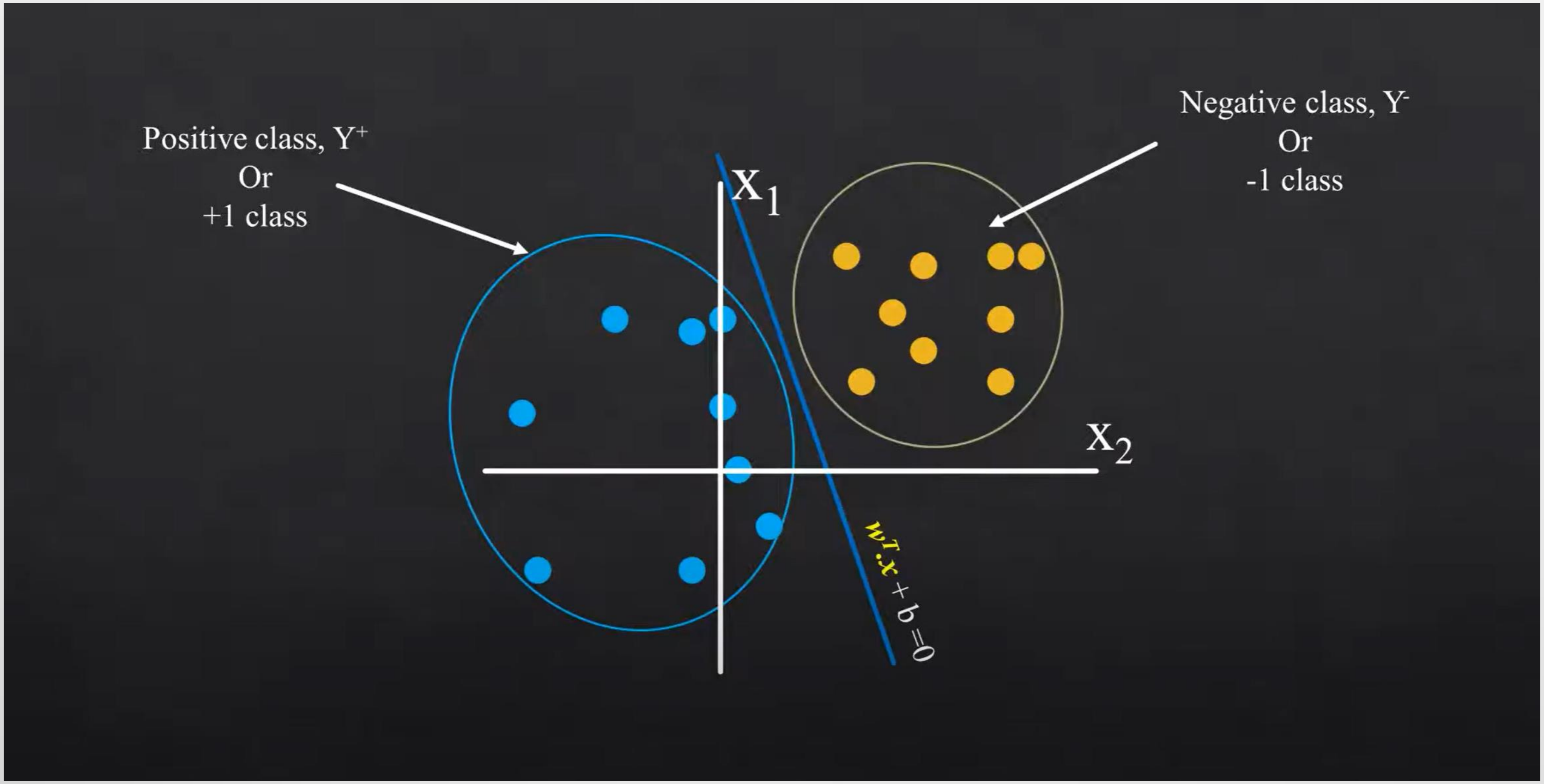
solve, $w^T.x + b$

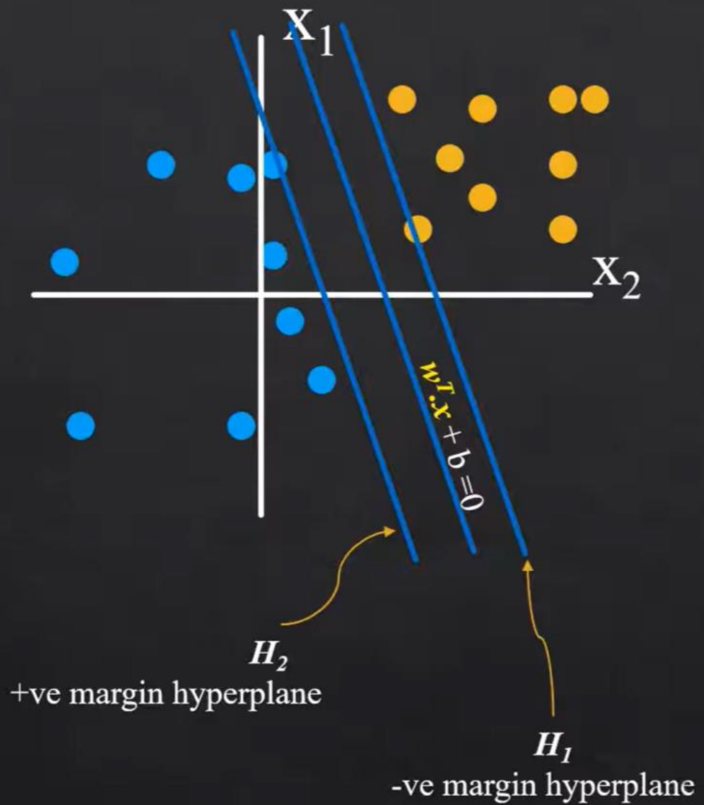
$$=(-4\ -1)*\begin{pmatrix}0.2\\2\end{pmatrix} + 4$$


$w = \begin{pmatrix} -4 \\ -1 \end{pmatrix}$, b = 4
X_1
X_2
(0,4)
(0.2 , 2)
(1.1 , 1)
(1,0)
$w^T.x$ + b =0
$x = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1.1 \\ 1 \end{pmatrix}$
solve, $w^T.x$ + b
$=(-4\ -1)*\begin{pmatrix} 1.1 \\ 1 \end{pmatrix} + 4$
= -1.4
Negative value for a yellow data point
$x = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0.2 \\ 2 \end{pmatrix}$
solve, $w^T.x$ + b
$=(-4\ -1)*\begin{pmatrix} 0.2 \\ 2 \end{pmatrix} + 4$
= +1.2
Positive value for a blue data point

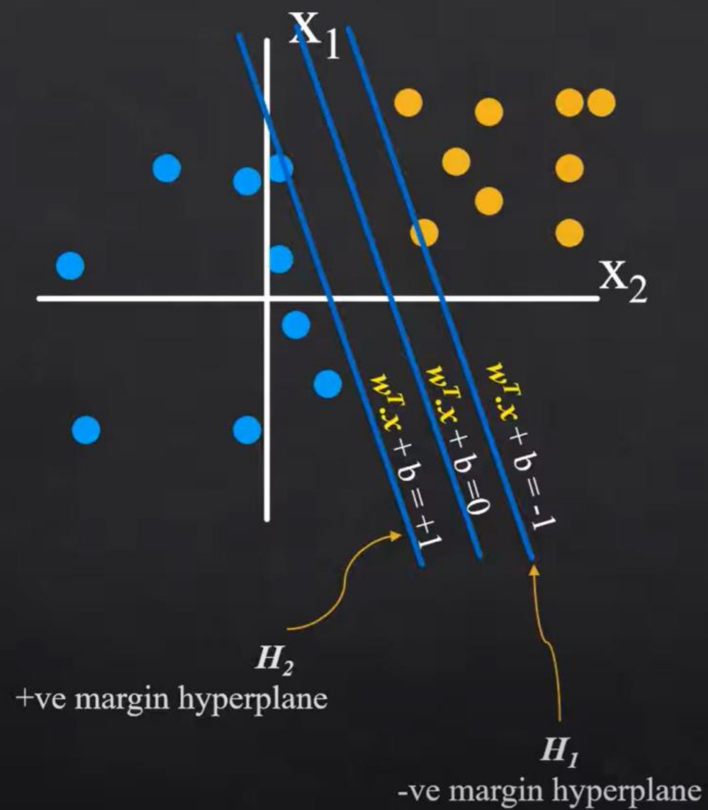


Positive class, Y^+
Or
+1 class
x_1
x_2

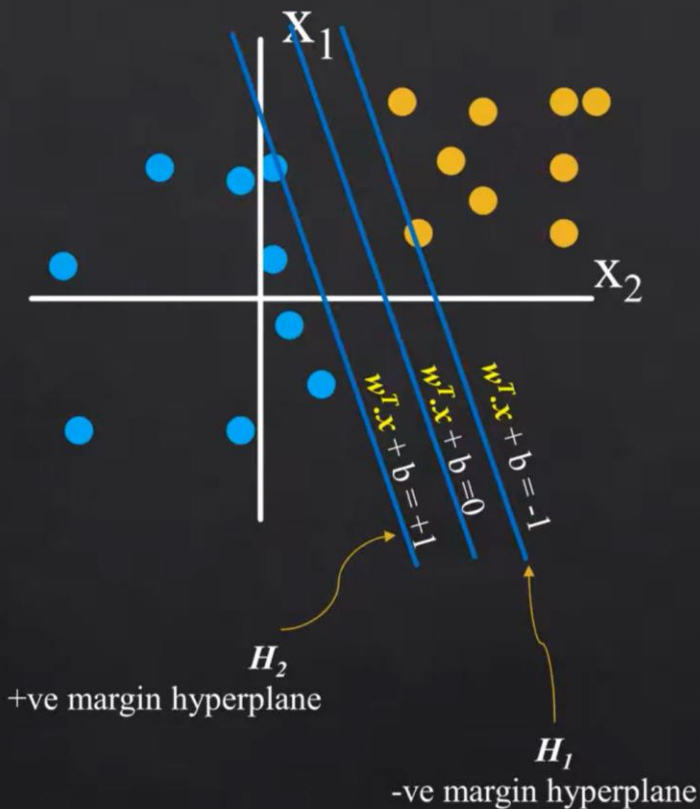

Positive class, Y+
Or
+1 class
Negative class, Y-
Or
-1 class
X1
X2
wT.x + b =0



In SVM,
We have two more hyperplanes passing through the support vectors.



In SVM,
We have two more hyperplanes passing through the support vectors.



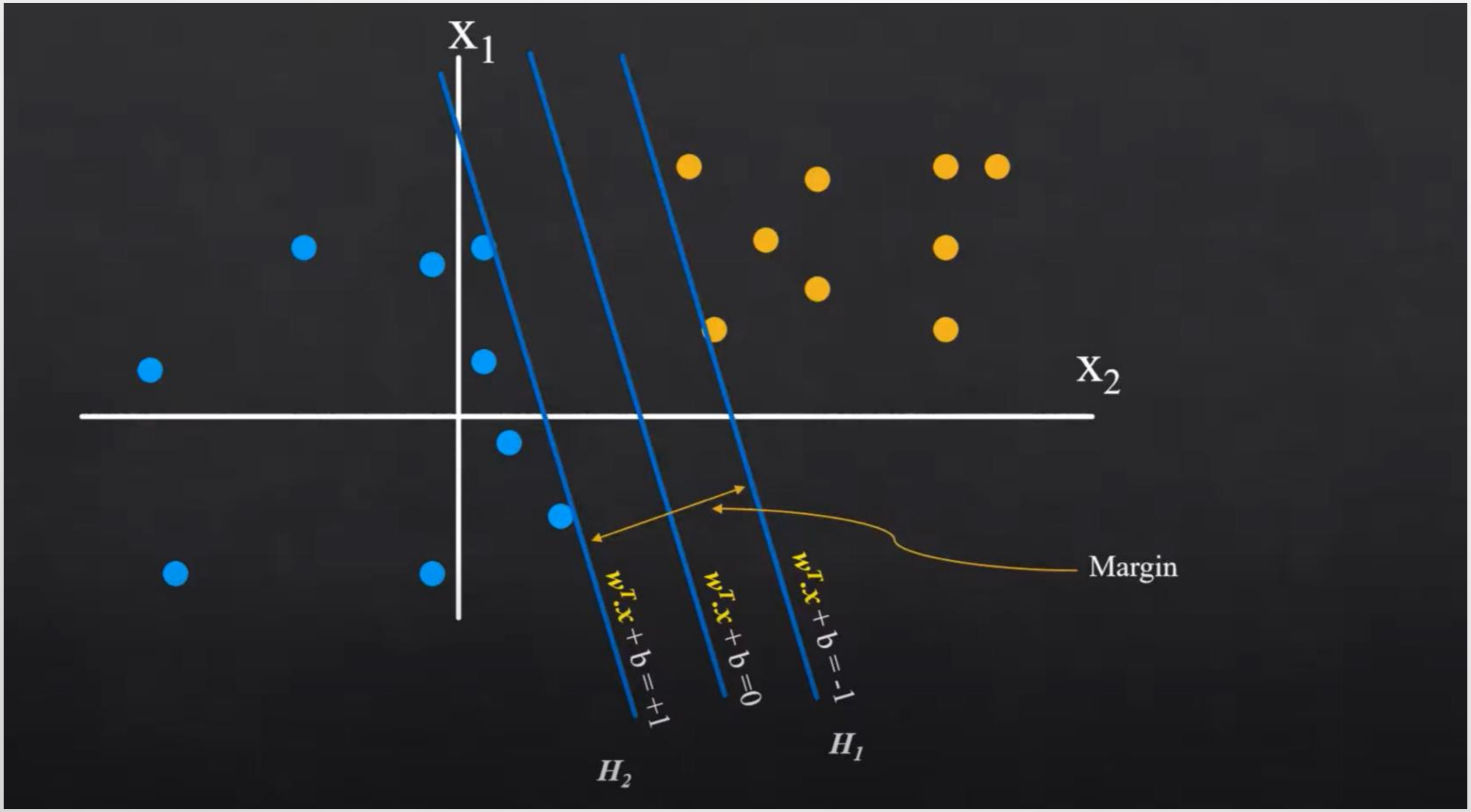
In SVM,
We have two more hyperplanes passing through the support vectors.

For each vector x_i

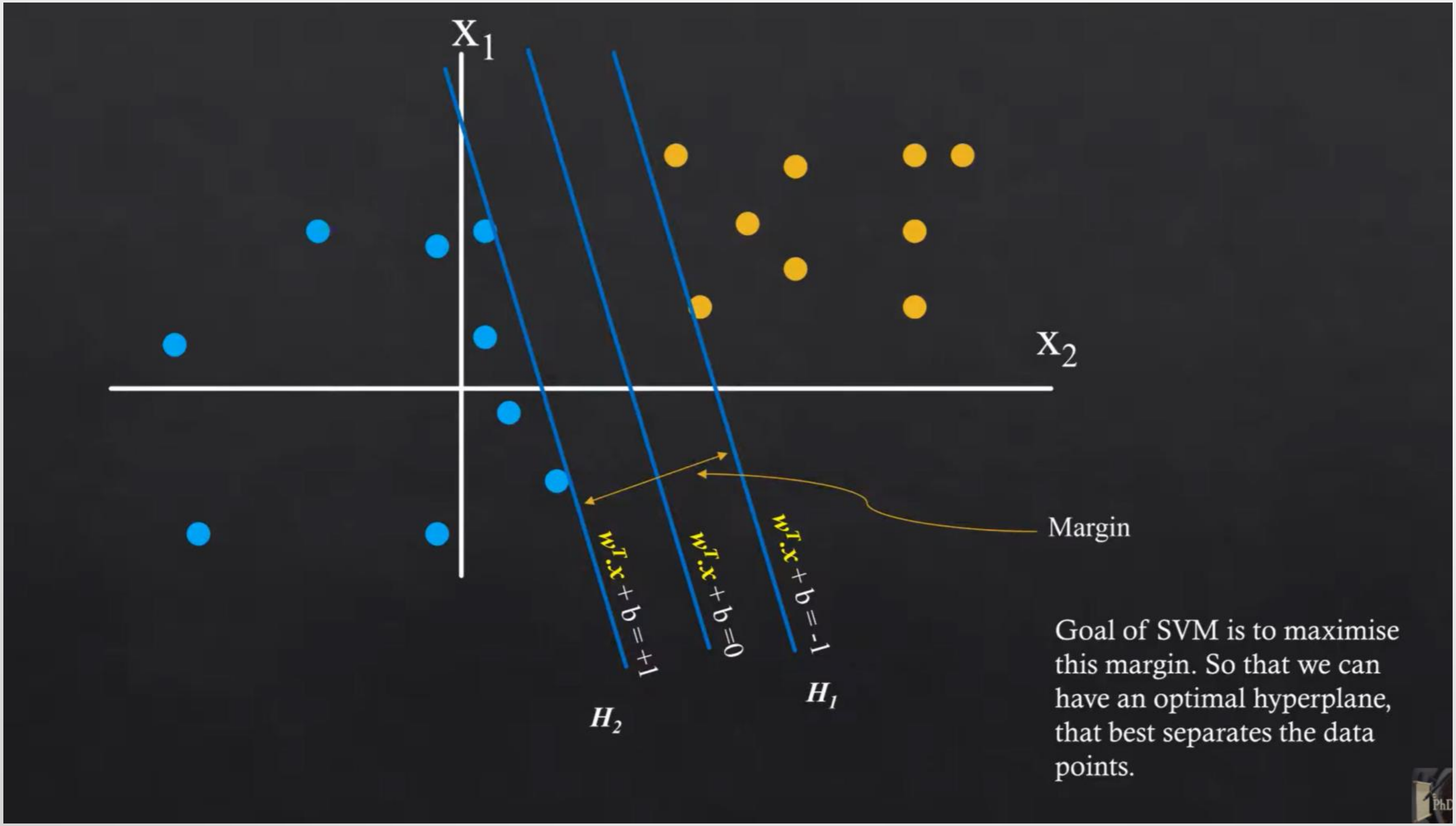
$w^T. x_i + b \geq +1$, for positive class, Y^+ (blue dots)

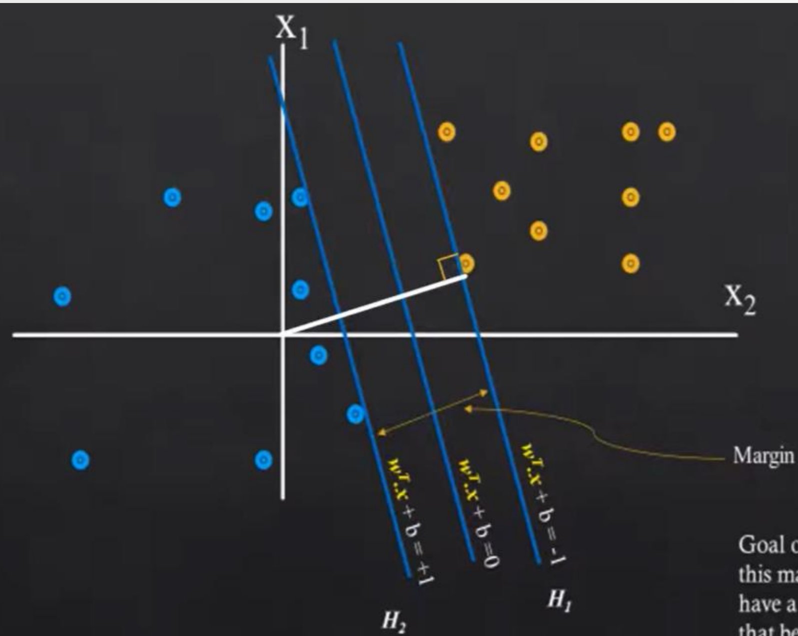
and

$w^T. x_i + b \leq -1$, for negative class, Y^- (yellow dots)



X1
X2
Margin
wT.x + b = +1
wT.x + b =0
wT.x + b = -1
H2
H1


X1
X2
$w^T.x + b = +1$
$w^T.x + b = 0$
$w^T.x + b = -1$
Margin
H2
H1
Goal of SVM is to maximise this margin. So that we can have an optimal hyperplane, that best separates the data points.



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

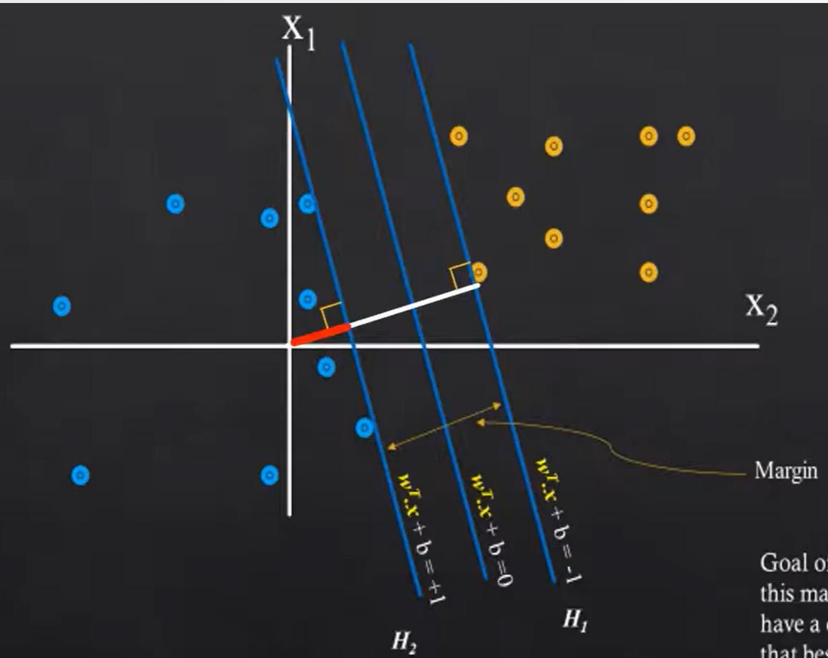
Margin = ?

$$H_1 : w^T . x_2 + b = -1$$

$$H_2 : w^T . x_1 + b = +1$$

Shortest distance between H_1 hyperplane and origin is

$$D_1 = \frac{(-1-b)}{\|w\|}$$



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

Margin = ?

$H_1 : w^T. x_2 + b = -1$

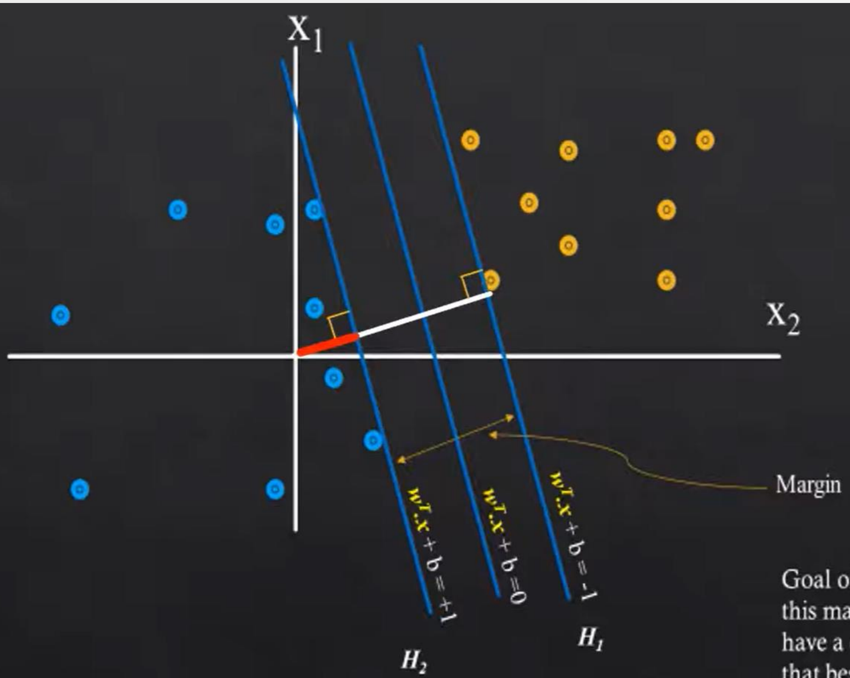
$H_2 : w^T. x_1 + b = +1$

Shortest distance between H_1 hyperplane and origin is

$$D_1 = \frac{(-1-b)}{\|w\|}$$

Shortest distance between H_2 hyperplane and origin is

$$D_2 = \frac{(1-b)}{\|w\|}$$



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

Margin = ?

$H_1 : w^T. x_2 + \mathrm{b} = -1$

$H_2 : w^T. x_1 + \mathrm{b} = +1$

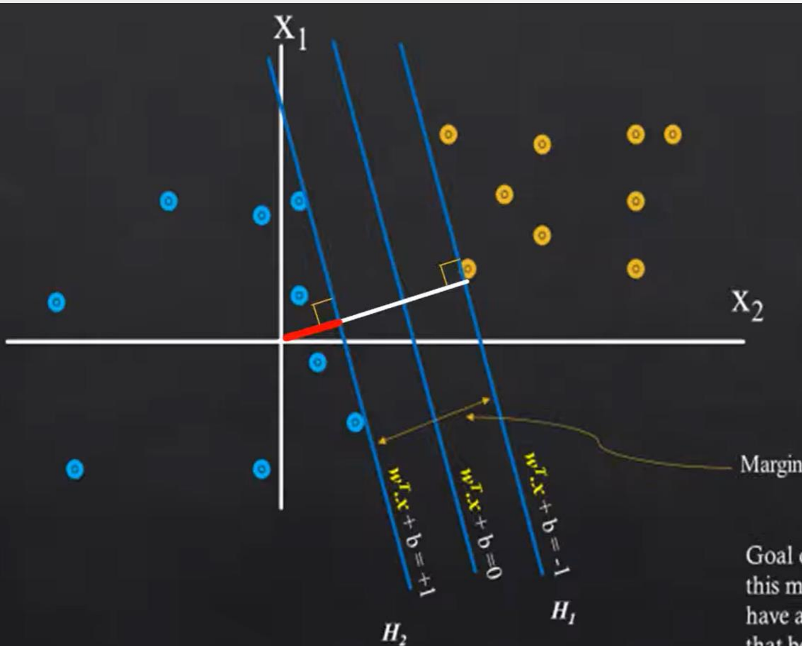
Shortest distance between H_1 hyperplane and origin is

$$D_1 = \frac{(-1-b)}{\|w\|}$$

Shortest distance between H_2 hyperplane and origin is

$$D_2 = \frac{(1-b)}{\|w\|}$$

Margin = | D_1 - D_2 |



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

Margin = ?

$H_1 : w^T. x_2 + \mathrm{b} = -1$

$H_2 : w^T. x_1 + \mathrm{b} = +1$

Shortest distance between H_1 hyperplane and origin is

$$D_1 = \frac{(-1-b)}{||w||}$$

Shortest distance between H_2 hyperplane and origin is

$$D_2 = \frac{(1-b)}{||w||}$$

$$\text{Margin} = | \ D_1 - D_2 \ |$$

$$= \left|\frac{(-1-b)}{||w||} - \frac{(1-b)}{||w||}\right| = 2/\ ||w||$$

Now we need to maximize $\frac{2}{\|w\|}$

Or

minimize: $\frac{\|w\|}{2}$

Objective function: $\emptyset(w^*, b^*) = \min \frac{\|w\|}{2}$

$$= \min\left(\frac{1}{2}\sqrt{w^T.w}\right)$$

$$= \min\left(\frac{1}{2}w^T.w\right)$$

Now we need to maximize $\frac{2}{\|\boldsymbol{w}\|}$

Or

minimize: $\frac{\|\boldsymbol{w}\|}{\mathbf{2}}$

Objective function: $\emptyset(w^*, b^*) = \min \frac{\|\boldsymbol{w}\|}{\mathbf{2}}$

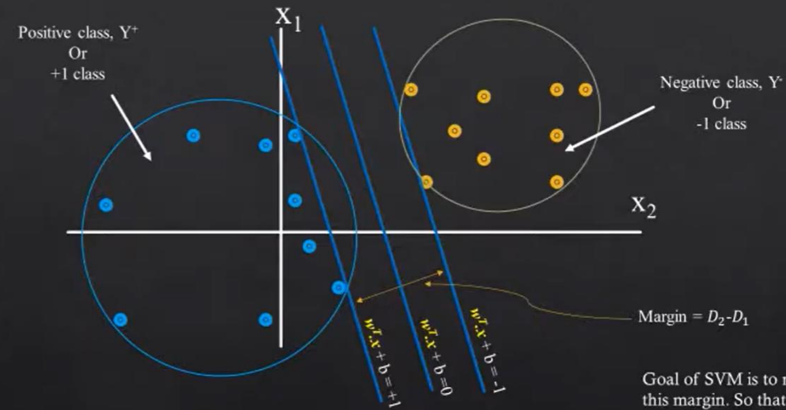
$= \min\left(\frac{1}{2}\sqrt{w^T.w}\right)$

$= \min\left(\frac{1}{2}w^T.w\right)$

S. t. $(\boldsymbol{w^T.x} + b)\,Y_i \geq 1$

Where, $Y_i = (Y^+, Y^-)$

Final Diagram



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

If $w^T.\,x + b \geq +1$, class Y^+

If $w^T.\,x + b \leq -1$, class Y^-

Rewrite as,

$(w^T.\,x + b)\,Y^+ \geq (+1)\,Y^+$

$(w^T.\,x + b)\,Y^- \leq (-1)\,Y^-$

We have, $Y^+ = +1$, and $Y^- = -1$

$$\overset{+ve}{(w^T.\,x + b)}\,\overset{+ve}{Y^+} \geq (+1)\,\overset{+ve}{Y^+}$$

$$\overset{+ve}{\overset{-ve}{(w^T.\,x + b)}\,\overset{-ve}{Y^-}} \leq (-1)\,\overset{-ve}{Y^-}$$

Final Diagram

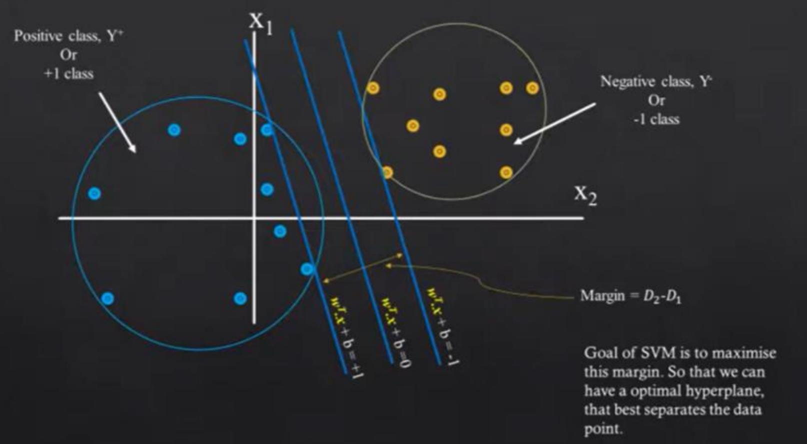
If $w^T . x + b \geq +1$, class Y^+

If $w^T . x + b \leq -1$, class Y^-

Rewrite as,

$(w^T . x + b)\, Y^+ \geq (+1)\, Y^+$

$(w^T . x + b)\, Y^- \leq (-1)\, Y^-$



We have, $Y^+ = +1$, and $Y^- = -1$

$$\overset{+ve}{(w^T . x + b)}\,\overset{+ve}{Y^+} \geq (+1)\,\overset{+ve}{Y^+}$$

$$\overset{+ve}{\overset{-ve}{(w^T . x + b)}\,\overset{-ve}{Y^-}} \leq (-1)\,\overset{-ve}{Y^-}$$

$$\Downarrow$$

$$(w^T . x + b)\, Y^- \geq (-1)(-1)$$

Final Diagram

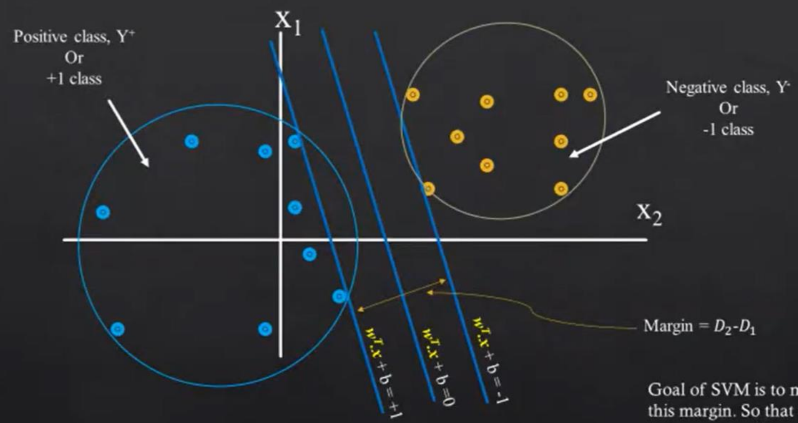
If $w^T . x + b \geq +1$, class Y^+

If $w^T . x + b \leq -1$, class Y^-

Rewrite as,

$(w^T . x + b)\ Y^+ \geq (+1)\ Y^+$

$(w^T . x + b)\ Y^- \leq (-1)\ Y^-$



Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.

We have, $Y^+ = +1$, and $Y^- = -1$

Therefore,

$(w^T . x + b)\ Y^+ \geq (+1)\ (+1)$

$(w^T . x + b)\ Y^- \leq (-1)\ (-1)$

$$(w^T . x + b)\ Y_i \geq 1$$

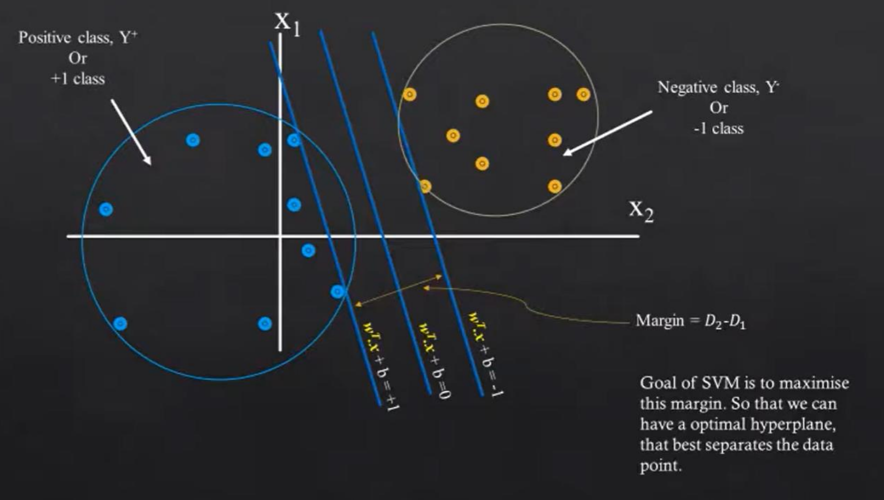
Where, $Y_i = (Y^+ , Y^-)$

This equation will classify the data points

Final Diagram

$$(w^T. x + \mathrm{b})\, Y_i \geq 1$$

Where, $Y_i = (Y^+, Y^-)$



For a data point (x)

if: $(w^T. x + \mathrm{b})\, Y_i \geq 1$

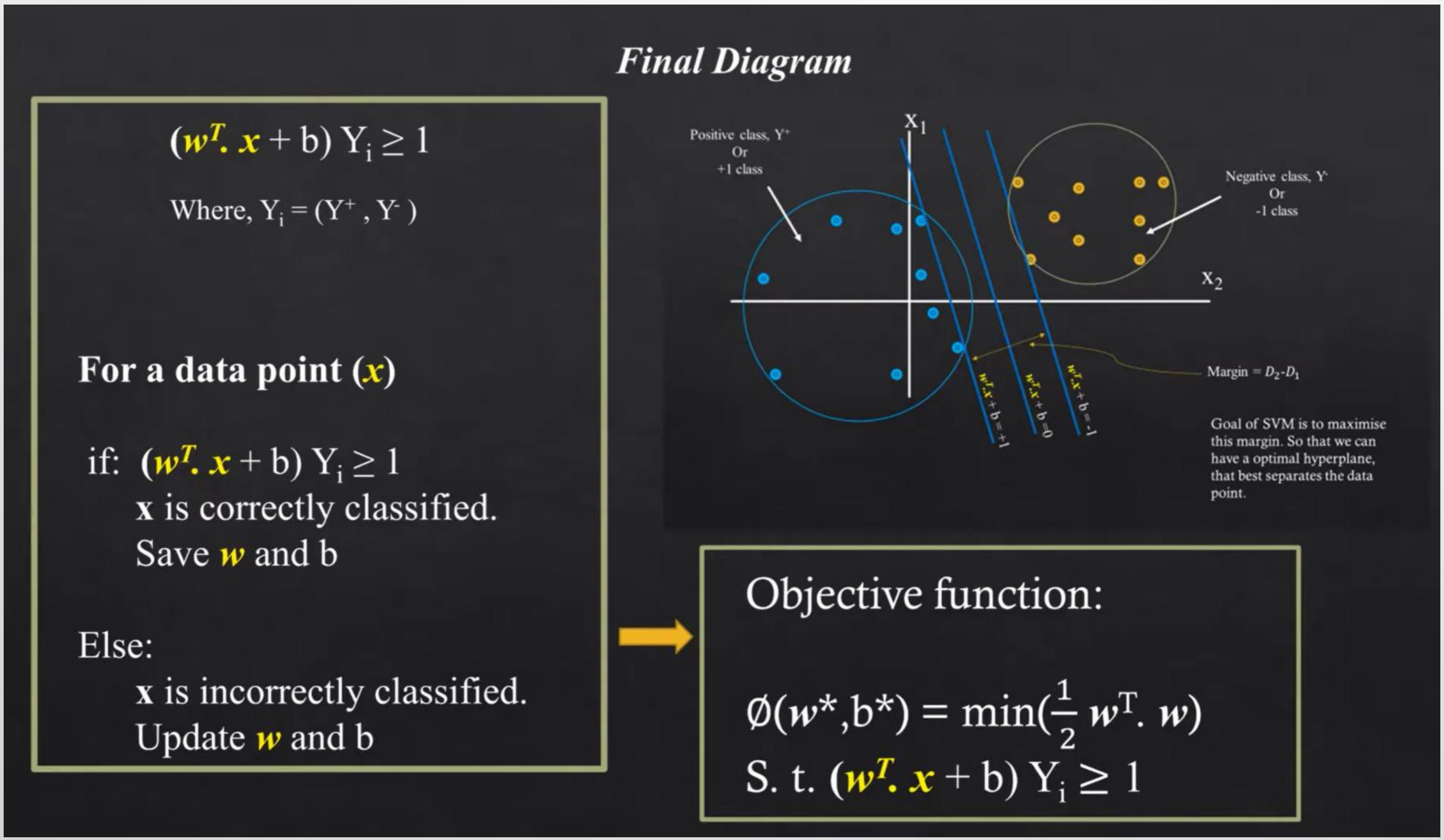
x is correctly classified.

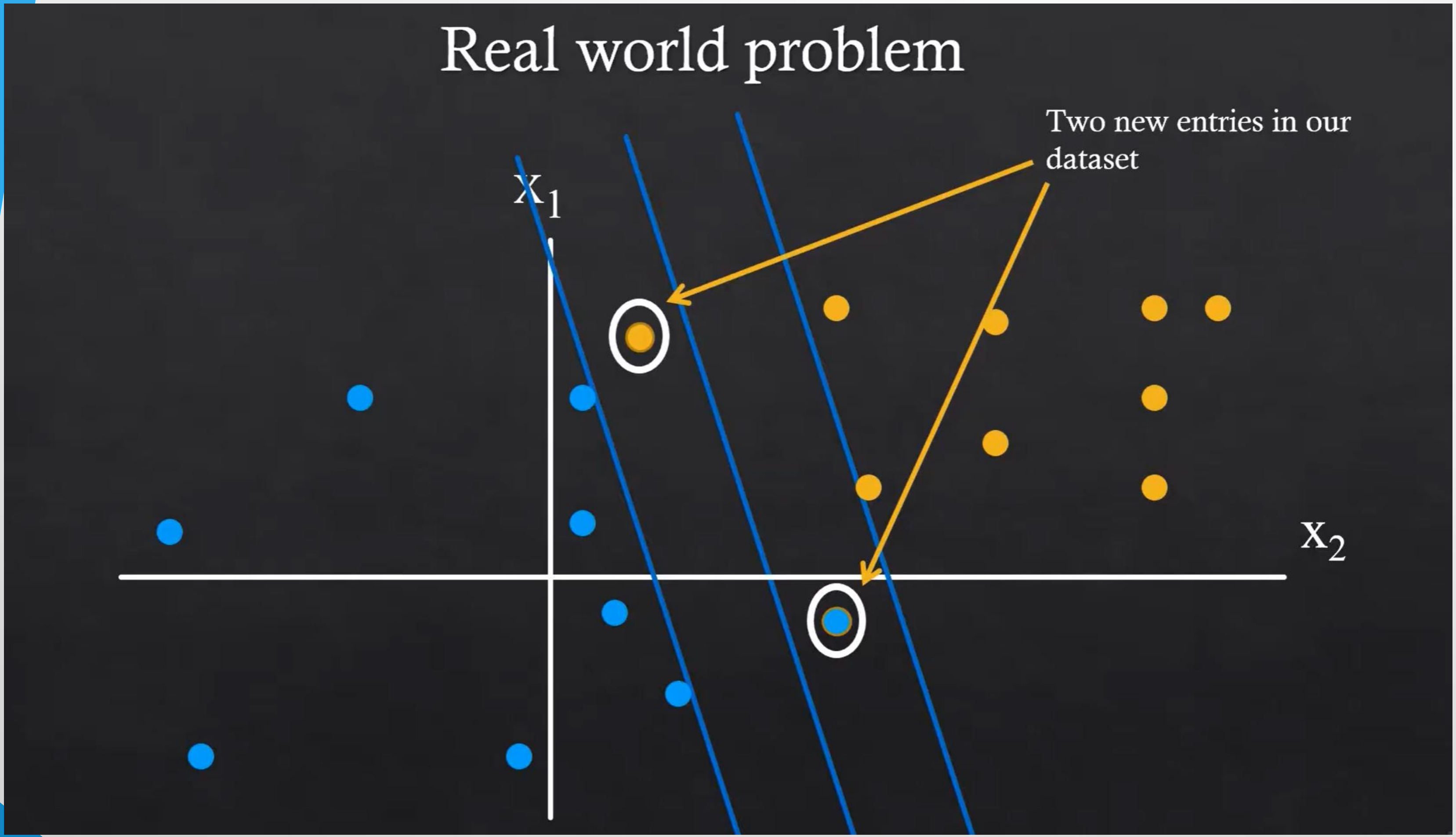
Save w and b

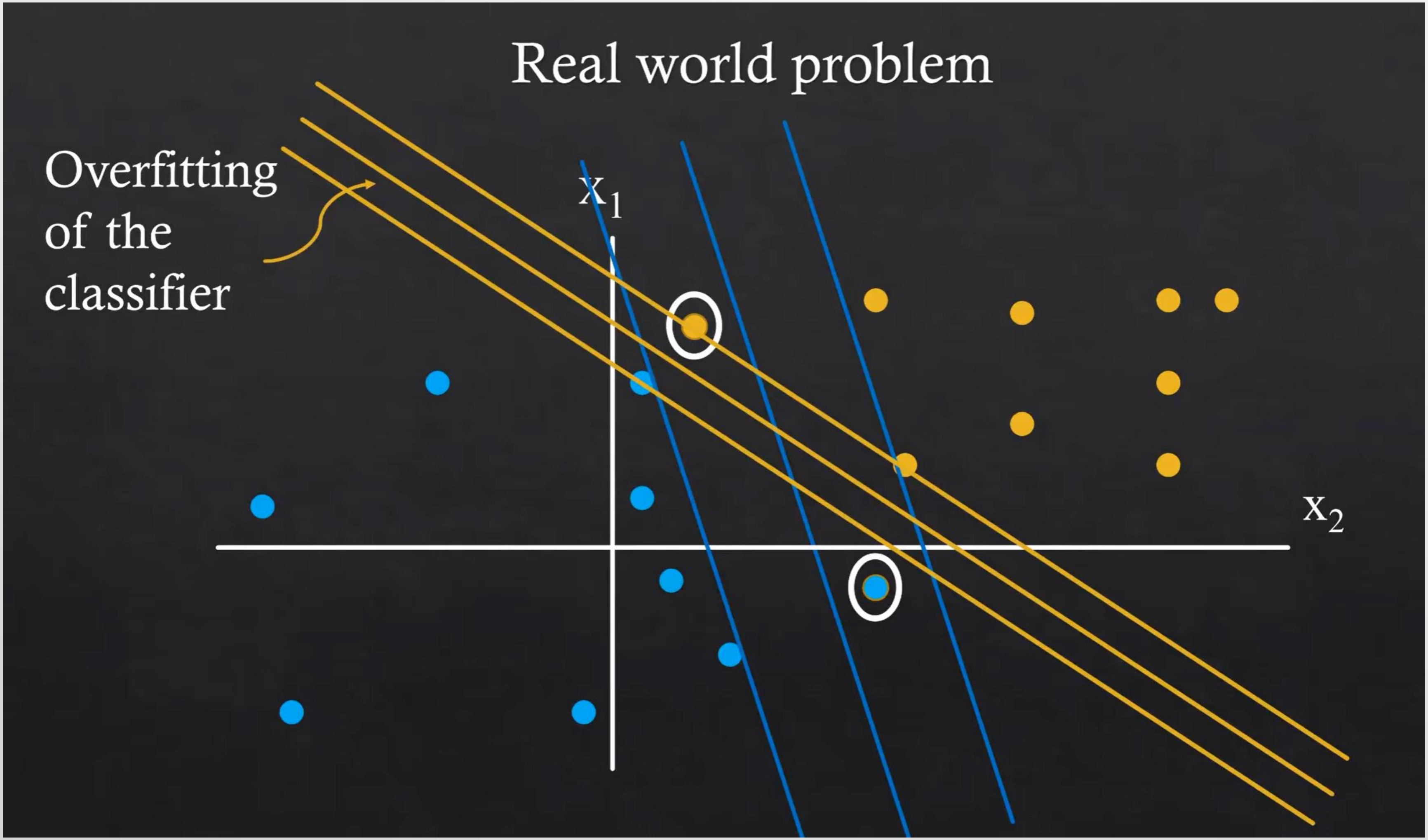
Else:

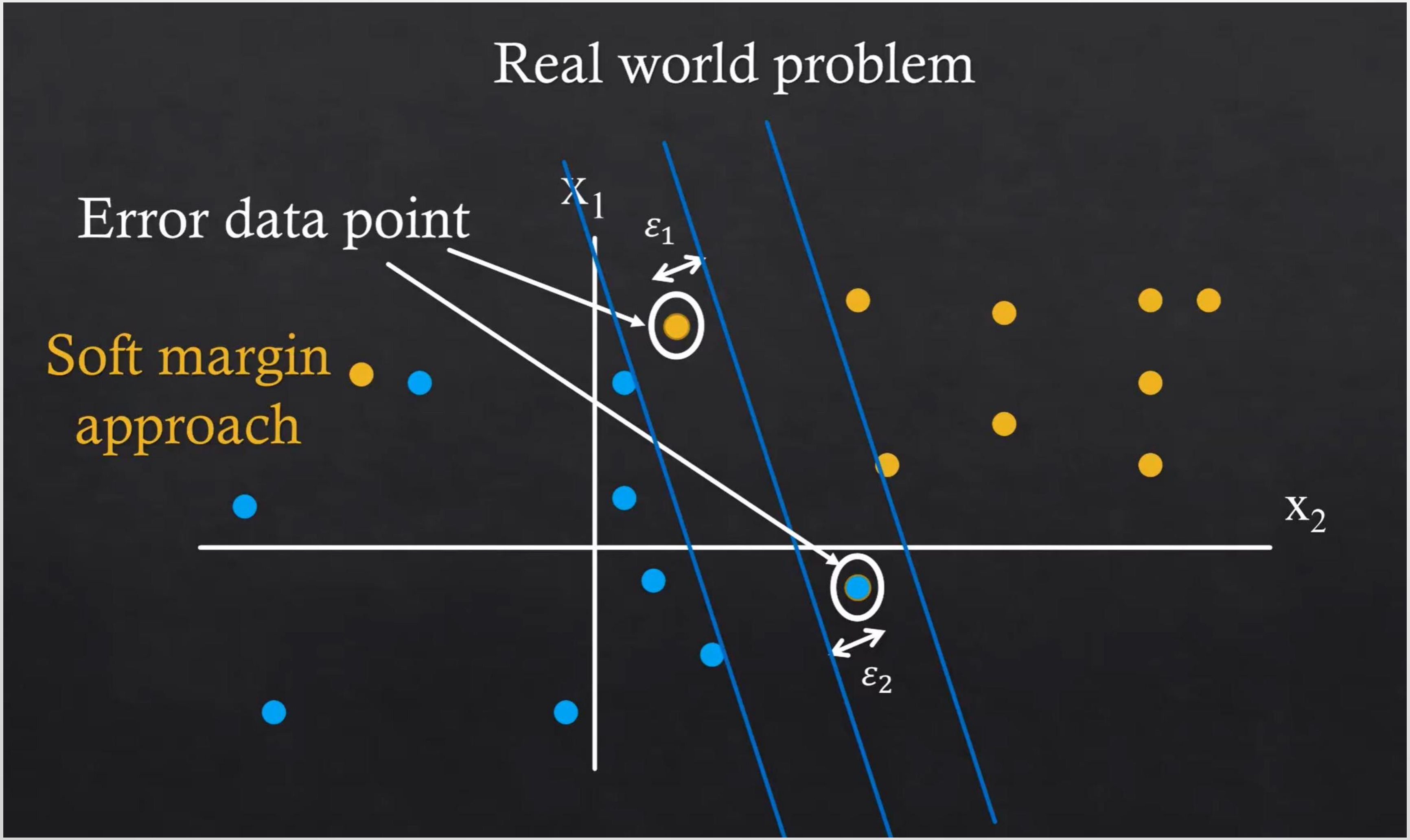
x is incorrectly classified.

Update w and b


Final Diagram
$(w^T . x + b)\ Y_i \geq 1$
Where, $Y_i = (Y^+ , Y^-)$
For a data point (x)
if: $(w^T . x + b)\ Y_i \geq 1$
x is correctly classified.
Save w and b
Else:
x is incorrectly classified.
Update w and b
Positive class, Y^+ Or +1 class
x_1
x_2
Negative class, Y^- Or -1 class
$w^T.x + b = +1$
$w^T.x + b = 0$
$w^T.x + b = -1$
Margin = D_2-D_1
Goal of SVM is to maximise this margin. So that we can have a optimal hyperplane, that best separates the data point.
Objective function:
$\emptyset(w^*, b^*) = \min(\frac{1}{2} w^T . w)$
S. t. $(w^T . x + b)\ Y_i \geq 1$


Real world problem
Two new entries in our dataset
x_1
x_2


Real world problem
Overfitting
of the
classifier
x_1
x_2


Real world problem
Error data point
Soft margin
approach
x_1
x_2
ε_1
ε_2

The idea here is to *allow SVM to make a certain number of mistakes and keep margin as wide as possible so that other points can still be classified correctly*. This can be done simply by modifying the objective function of SVM.