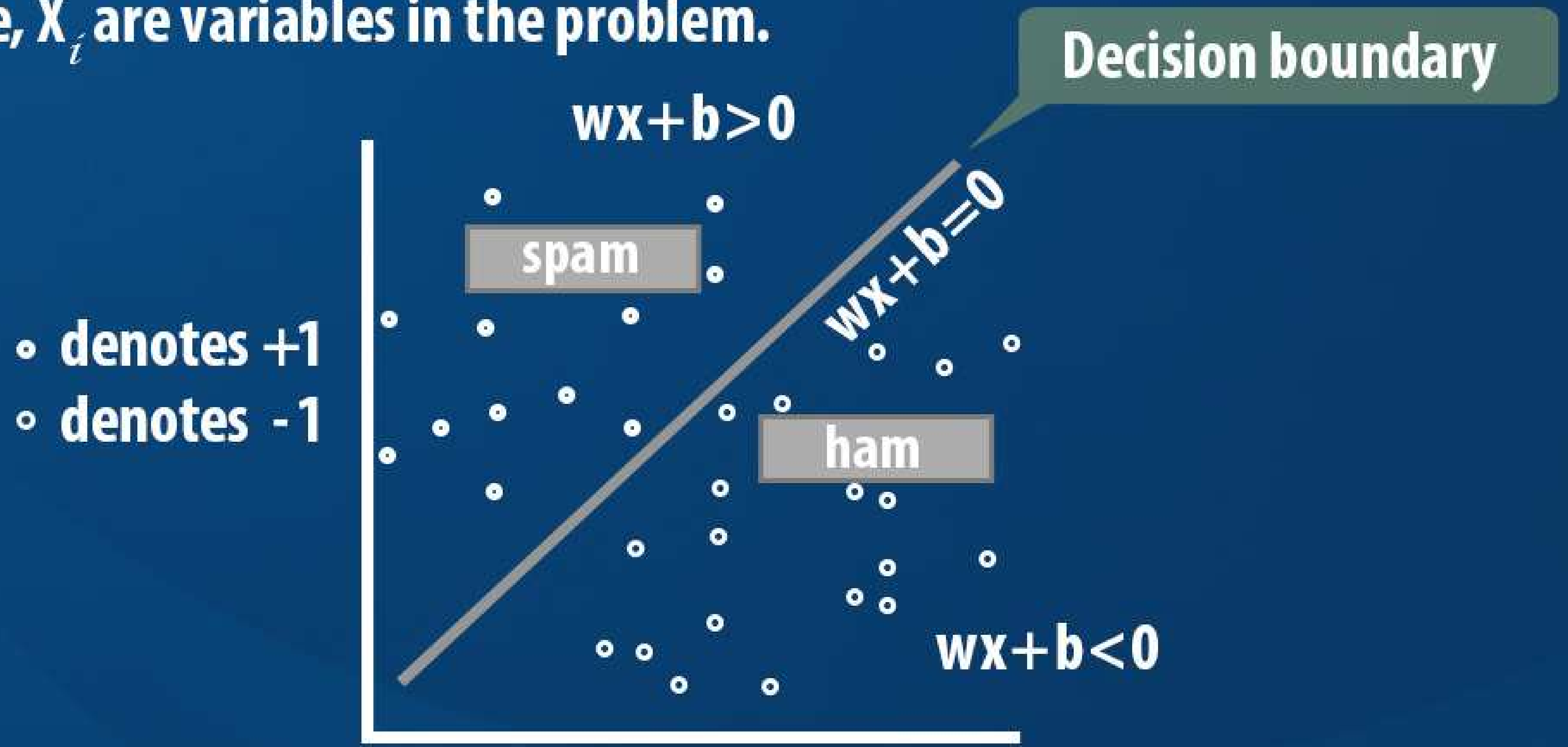
SUPPORT VECTOR MACHINES FOR LINEARLY SEPARABLE PROBLEMS

Consider an application of predicting an incoming email as Spam(+1) and Ham(-1). It is a binary classification problem where, objective is to learn f(x) such that:

$$f(x) = \begin{cases} +1, & \text{if } W_1X_1 + W_2X_2 + \ldots\ldots + W_n X_n \geq \theta \\ -1, & \text{otherwise} \end{cases} \qquad (1)$$

where, X_i are variables in the problem.



$$W \cdot X + b = 0$$

$$W_0 + W_1X_1 + W_2X_2 = 0$$

$$W_0 + W \cdot X = 0$$

$$W_0 + \sum W_i X_i$$

$$(W_1 \cdot W_2)\ (X_1 \cdot X_2)$$

Figure 1: Support Vector Machines for Linearly separable problems

The key objective here is:

How to find the line or in other words, how to find W's?

HOW TO FIND THE BEST LINEAR SEPARABLE PROBLEMS?

There can exist several lines (defined by W's) as decision boundary, which line to select?

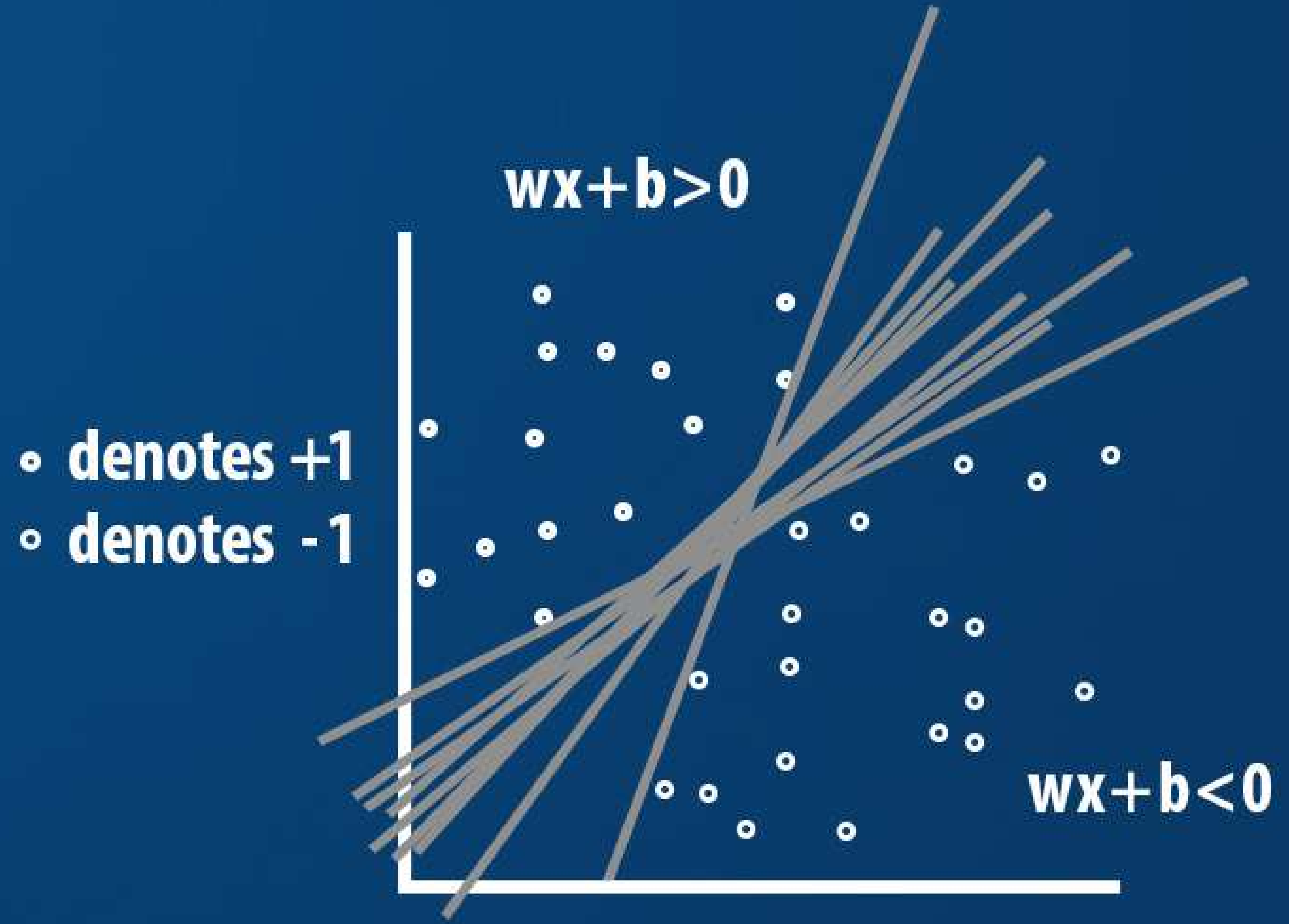


Figure 2: Multiple decision boundaries for some given binary classification problem

The best line is the one that has maximum margin.

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW

FINDING THE BEST DECISION BOUNDARY

Margin: Distance of closest examples from the decision line.

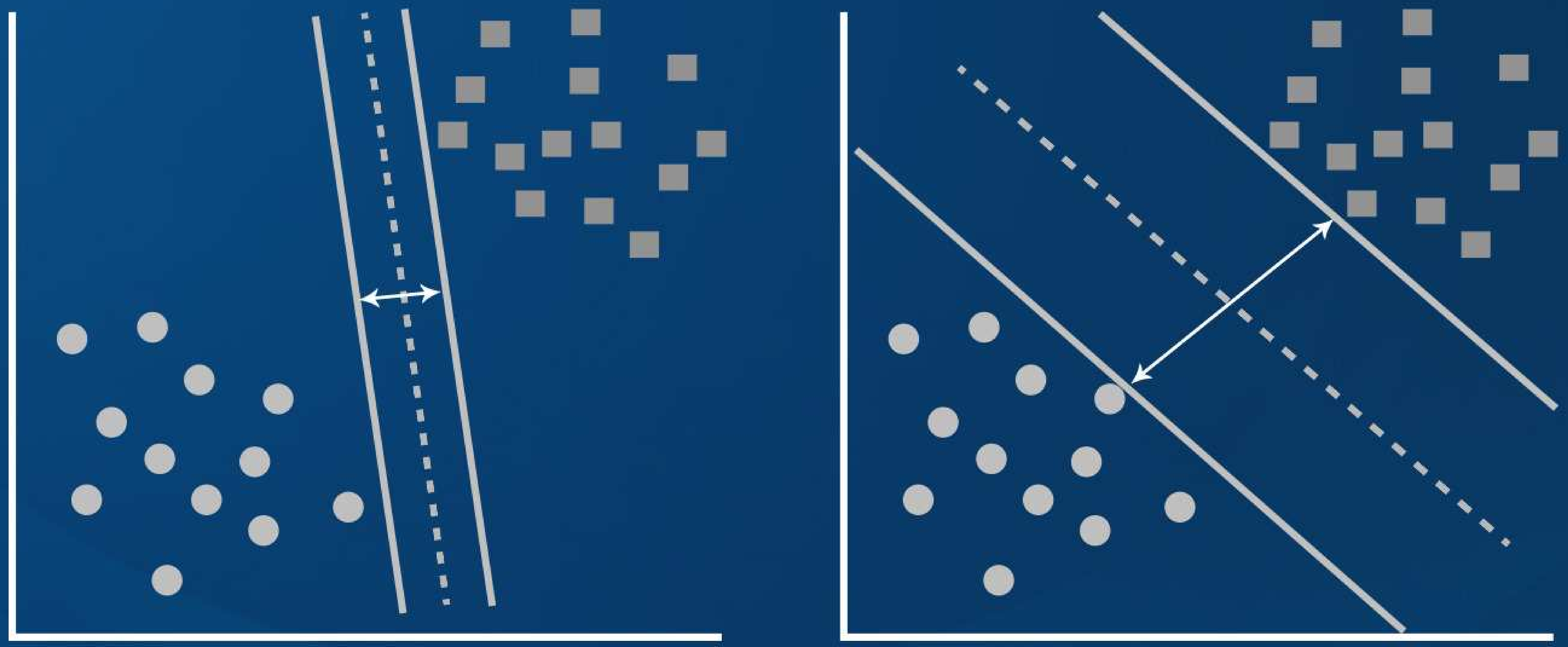


Figure 3: Finding best decision boundary based on margin.

DEFINITION OF SUPPORT VECTORS

The separating decision boundary is defined by support vectors.

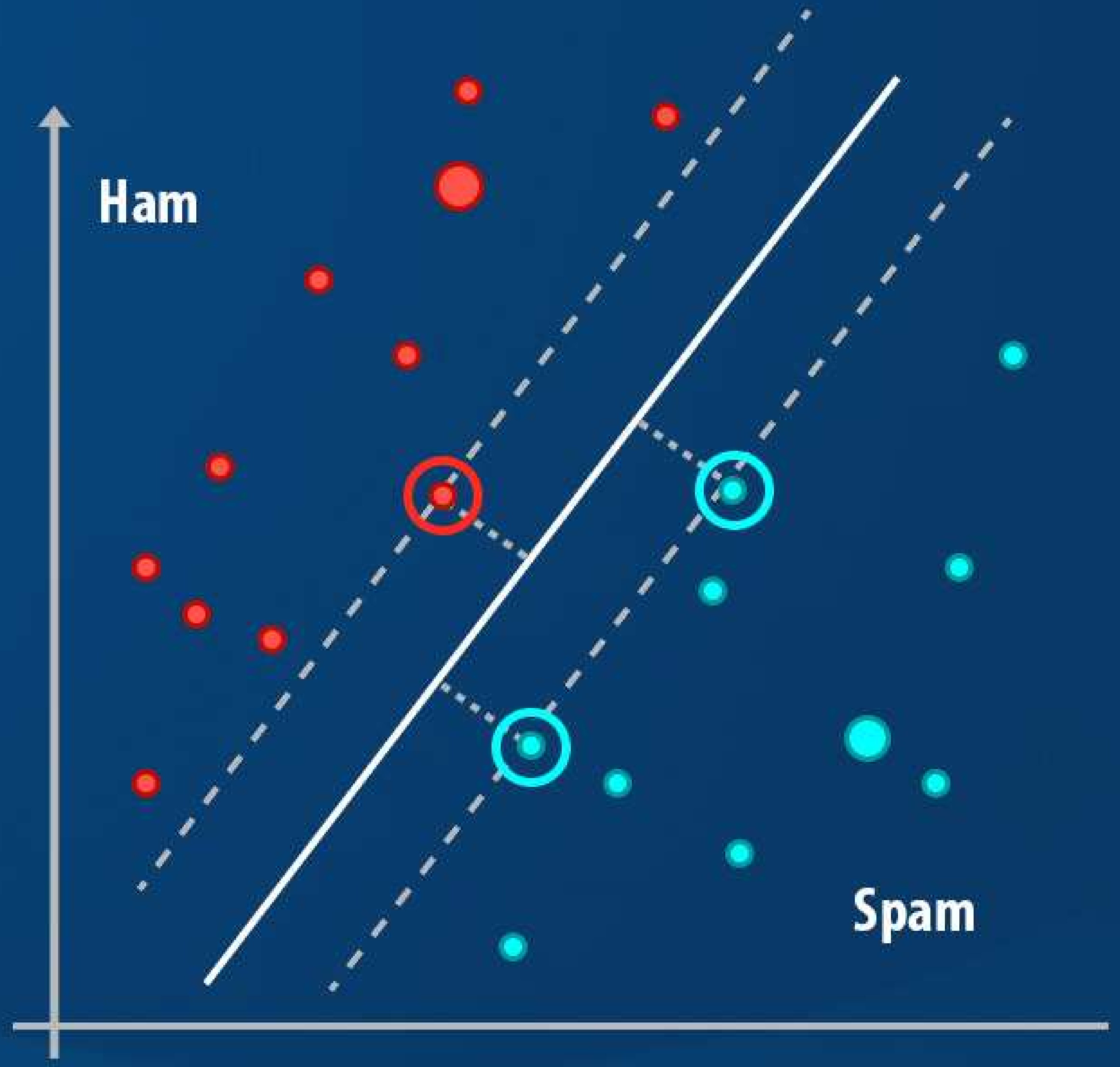
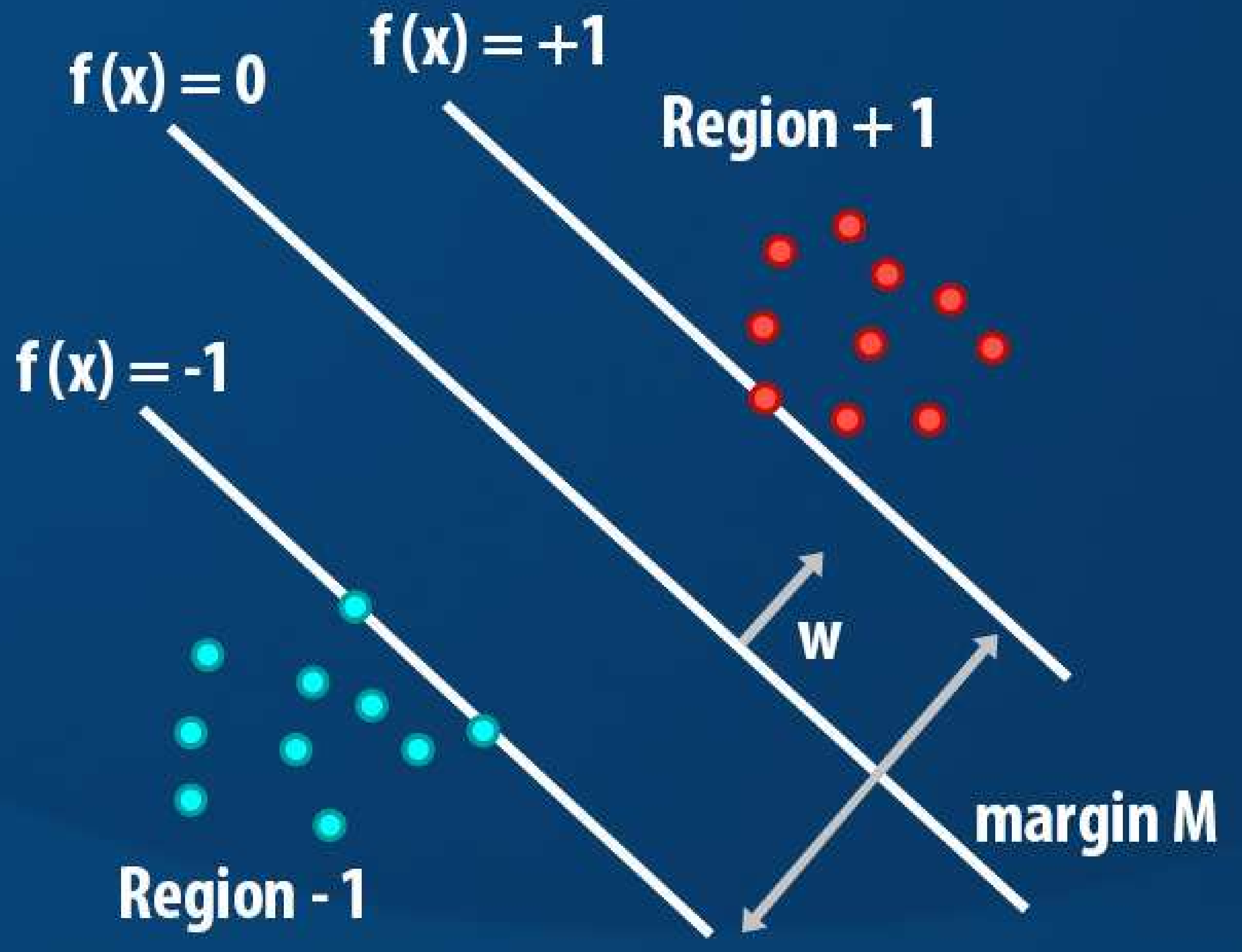


Figure 4: The highlighted data points are called Support Vectors

DEFINING MARGIN

1. Maximise Margin.
2. Define class +1 in one region, and -1 in some another region.
3. Make regions as far apart as possible.



$f(x) = w \cdot x + b$

$f(x) > +1$ is region +1

$f(x) < -1$ is region -1

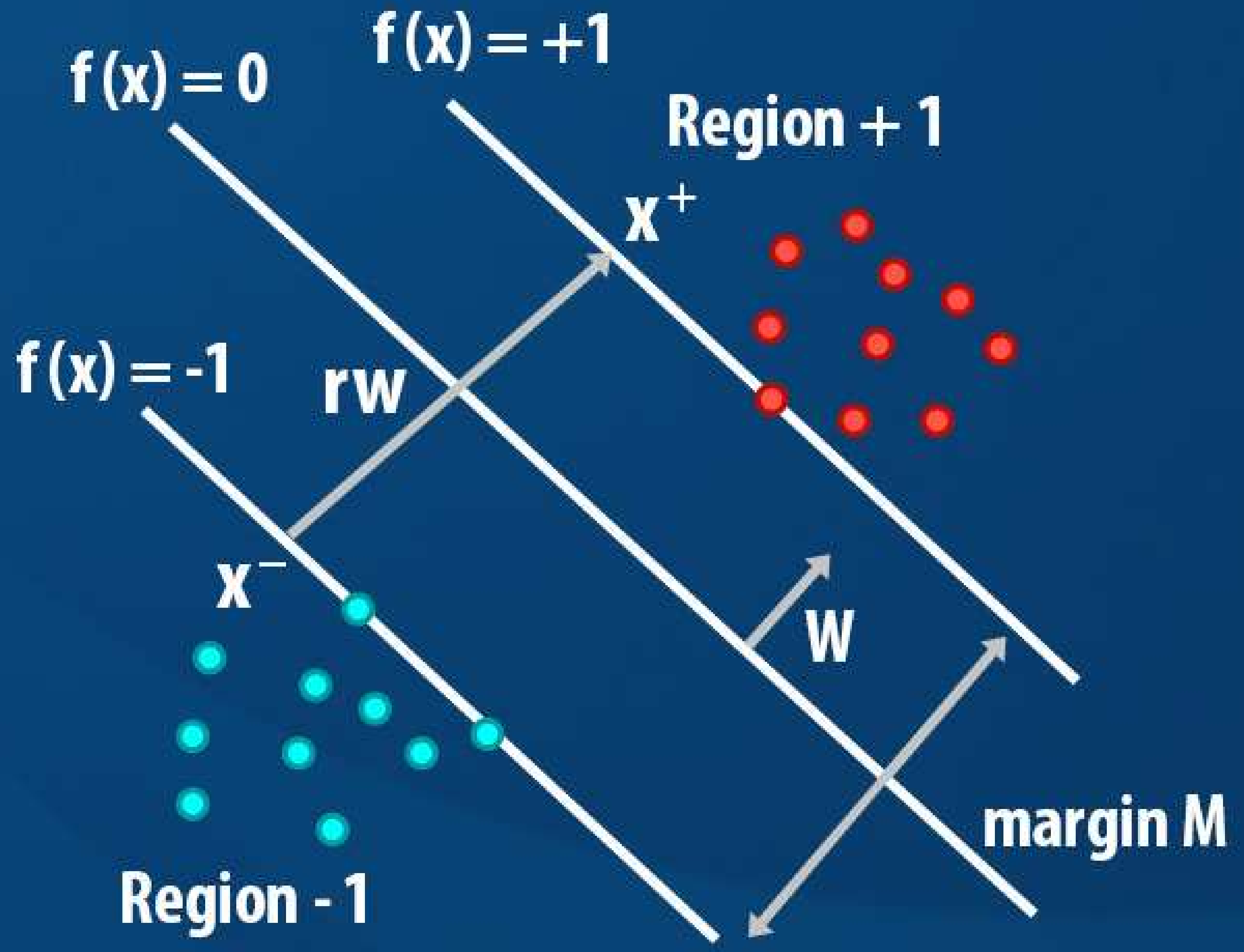
Figure 5: Defining functions for boundaries

HOW TO MAXIMIZE MARGIN?

1. The first thing to note here is that the w's are perpendicular to the decision boundary.
2. Choose two data points x^+ such that $f(x^+) = +1$ and x^- such that $f(x^-) = -1$.

Therefore,

$$x^+ = x^- + r \cdot w \qquad (2)$$



$$f(x)=wx+b$$

$$w \cdot x^- + b = -1$$

$$w \cdot x^+ + b = +1$$

$$w\,(x^- + rw\,) + b = +1$$

$$wx^- + r\,||w\,||^2 \;\; + b = +1$$

$$w=(1,3)$$

$$w.w=1^2+3^2=10$$

$$||w\,||^2 = (\sqrt{1^2+3^2})^2 = (\sqrt{10})^2=10$$

$$r\,||w\,||^2 - 1 = 1$$

$$r = \frac{2}{||w\,||^2}$$

$$M = ||x^+ - x^-\,||^2$$

$$M = ||rw\,||= \frac{2}{||w\,||^2} \cdot ||w\,||$$

$$M = \frac{2}{||w\,||}$$

$$M = \frac{2}{\sqrt{(w^T \cdot w)}}$$

MAXIMISING MARGIN IS A OPTIMISATION PROBLEM

1. Classify each data point correctly
2. Maximise the margin with subject to constraints as follows

$$M = \arg\max_{w} \frac{2}{\sqrt{(w^T \cdot w)}} \qquad (3)$$

subject to constraint in Equations below such that:

$$y^i = +1 \longrightarrow w \cdot x^i + b \geq 1 \qquad (4)$$

$$y^i = -1 \longrightarrow w \cdot x^i + b \leq 1 \qquad (5)$$

where Equation 4 and Equation 5 can written as:

$$y^i (w \cdot x^i + b) \geq +1 \qquad (6)$$

It is optimisation problem solved using Lagrange multipliers method.

AMITY UNIVERSITY ONLINE
CAREERS OF TOMORROW