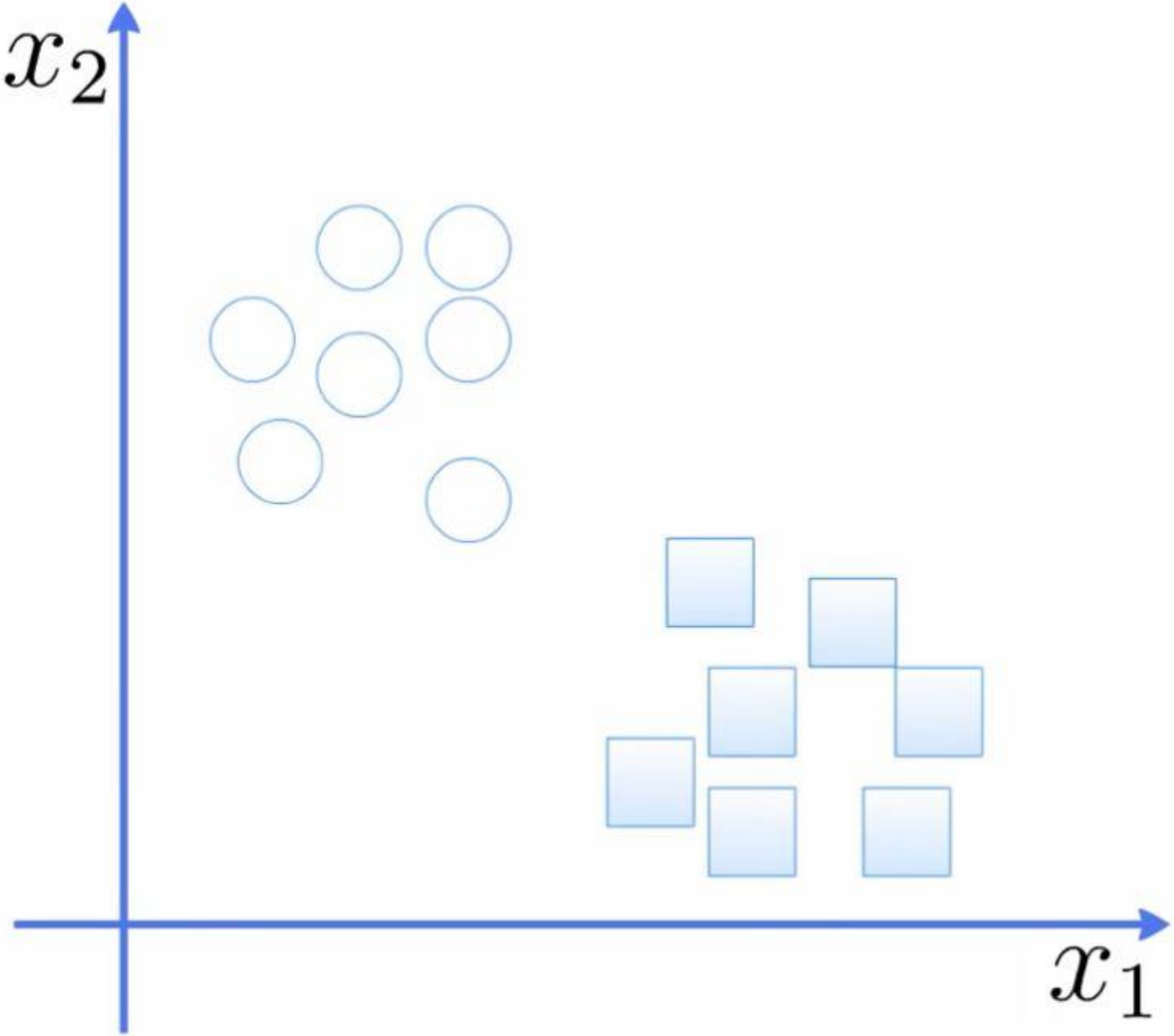
Support Vector Machine (SVM)

What it is?

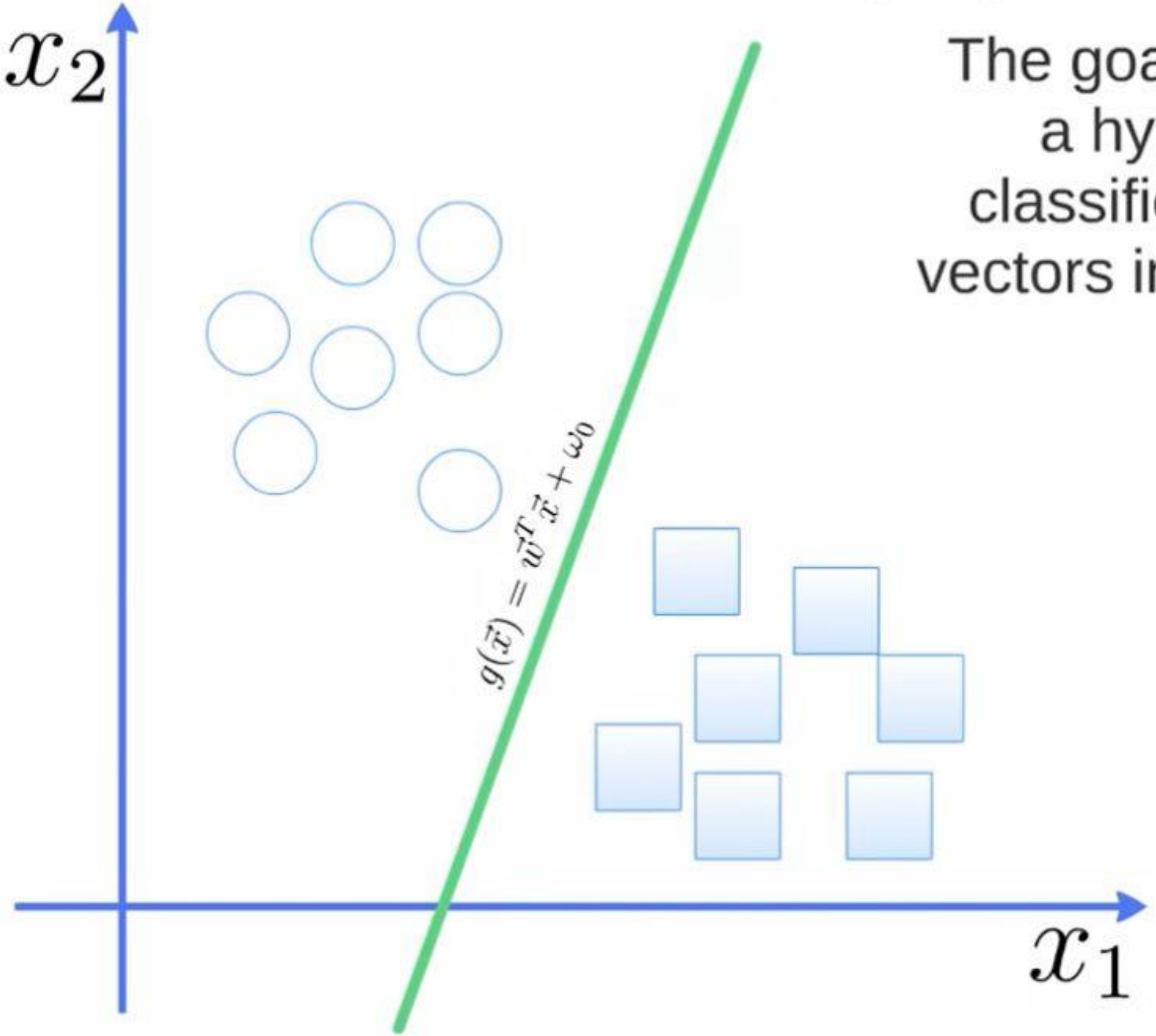
Numerical classifier that draws a single decision boundary that <u>maximizes the margin</u> between two classes of data

Classifier – Machine Learning Model

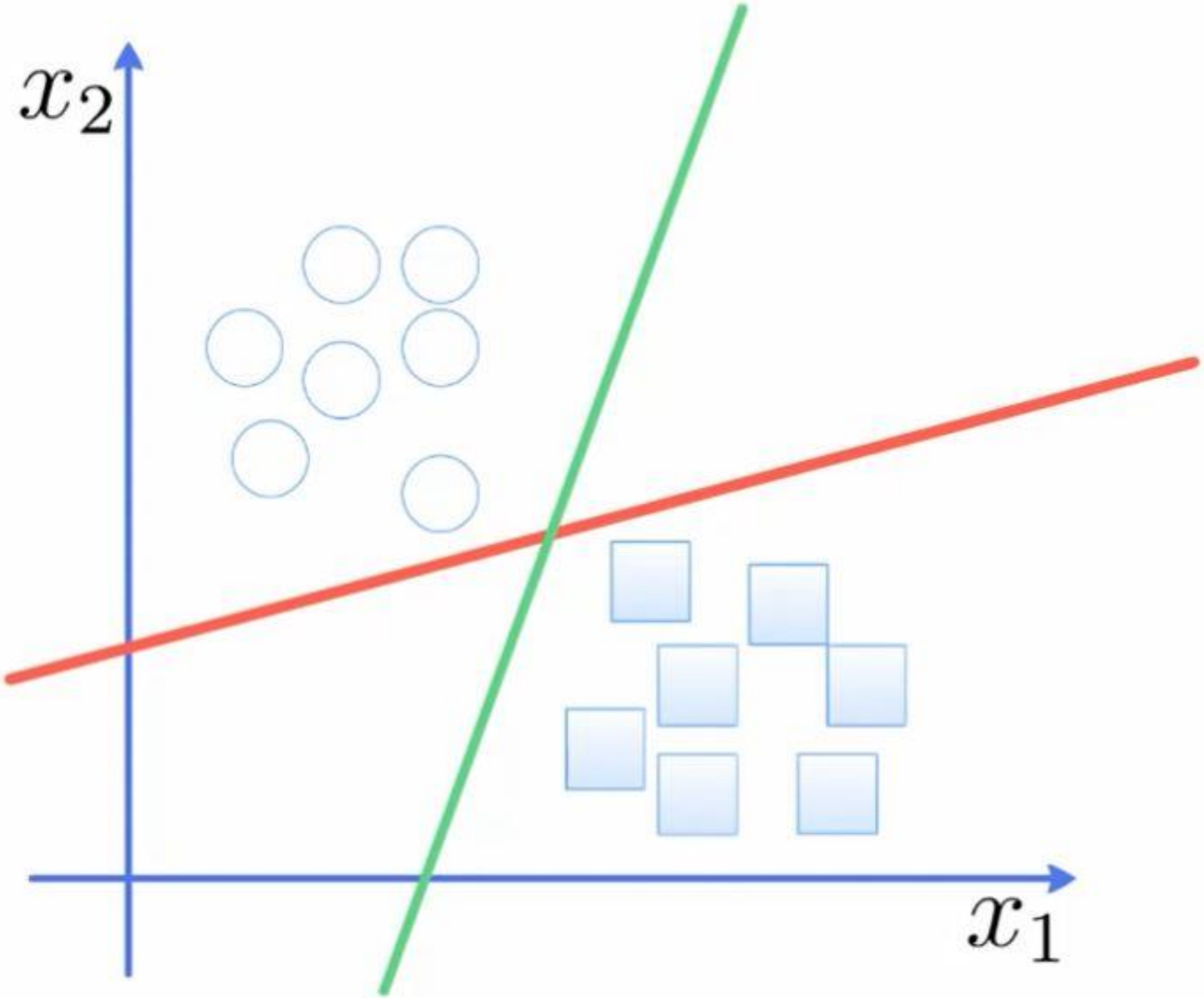
SVM for linearly separable binary sets



SVM for linearly separable binary sets

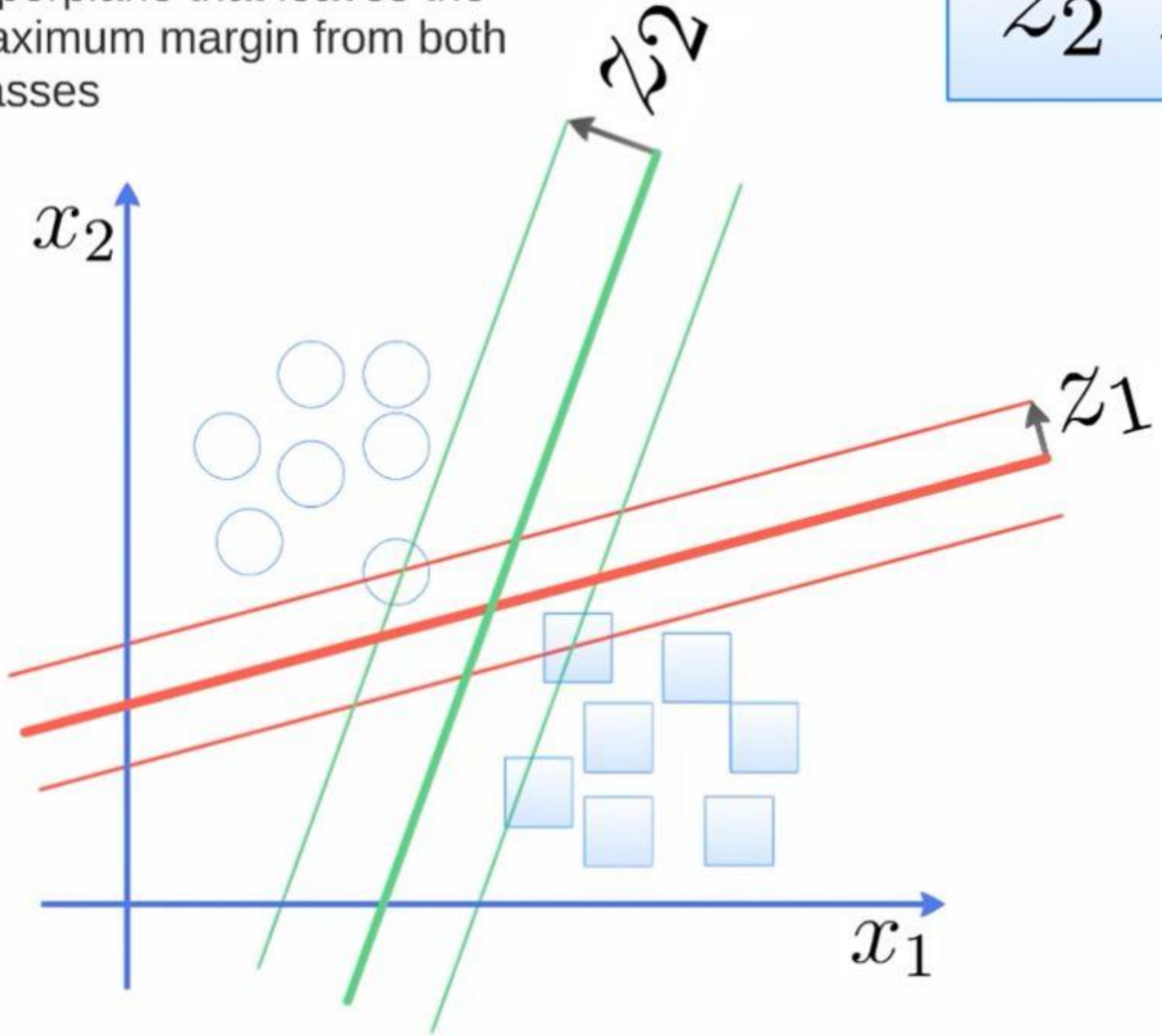


The goal is to design a hyperplane that classifies all training vectors in two classes


x_2
x_1

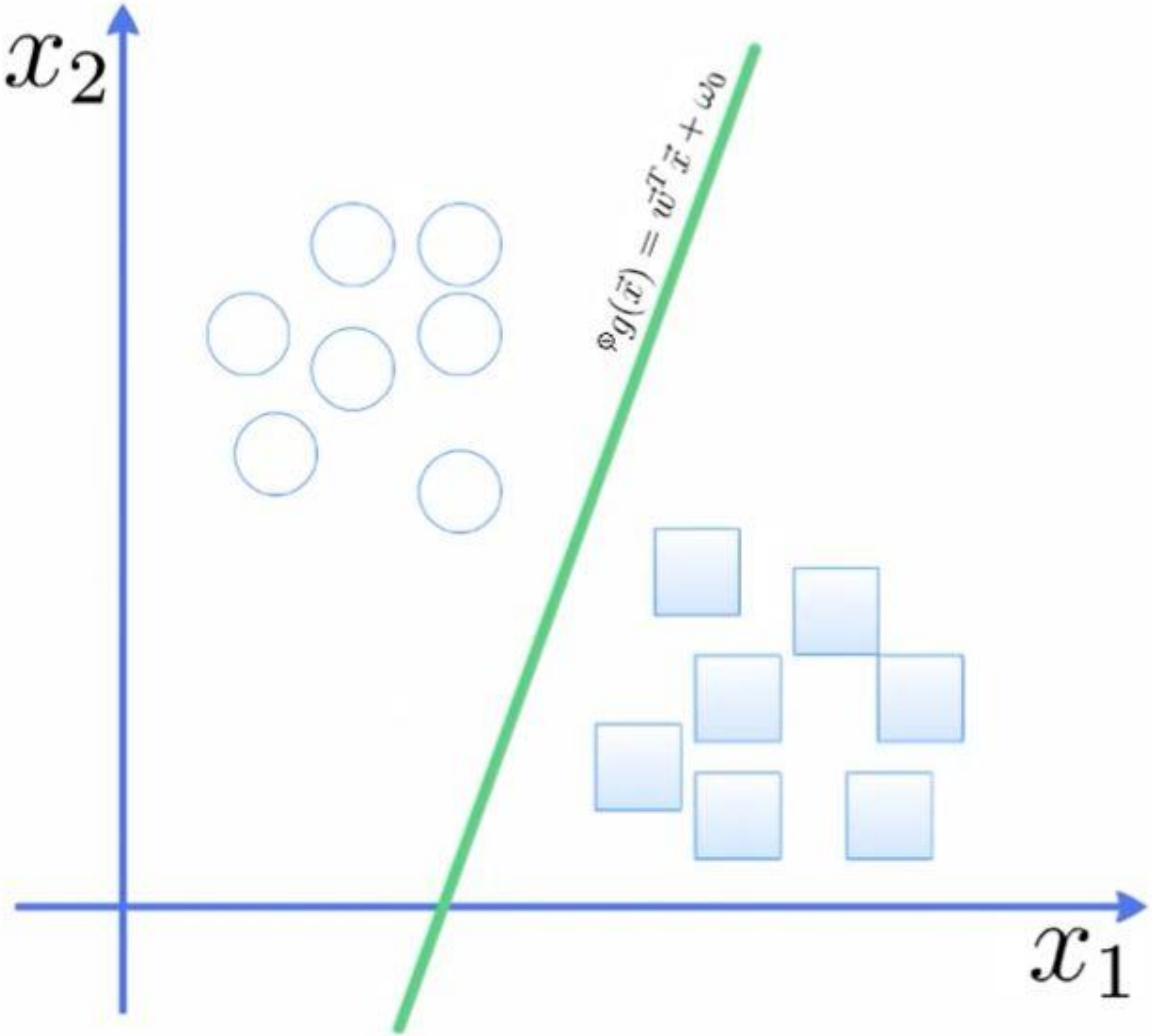
The best choice will be the hyperplane that leaves the maximum margin from both classes

$$z_2 > z_1$$



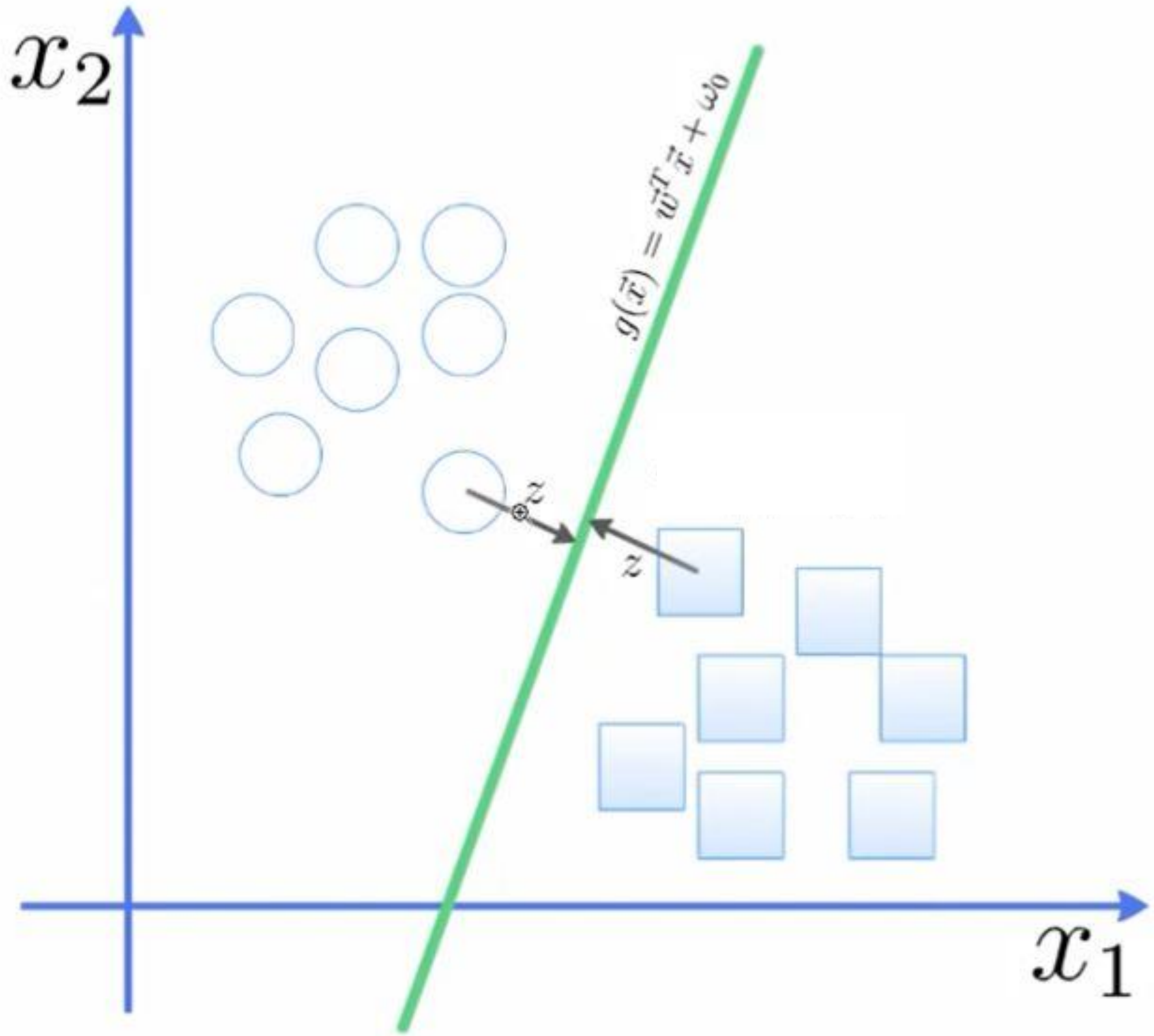
$$g(\vec{x}) \geqslant 1, \quad \forall \vec{x} \in \text{class } 1$$

$$g(\vec{x}) \leqslant -1, \quad \forall \vec{x} \in \text{class } 2$$

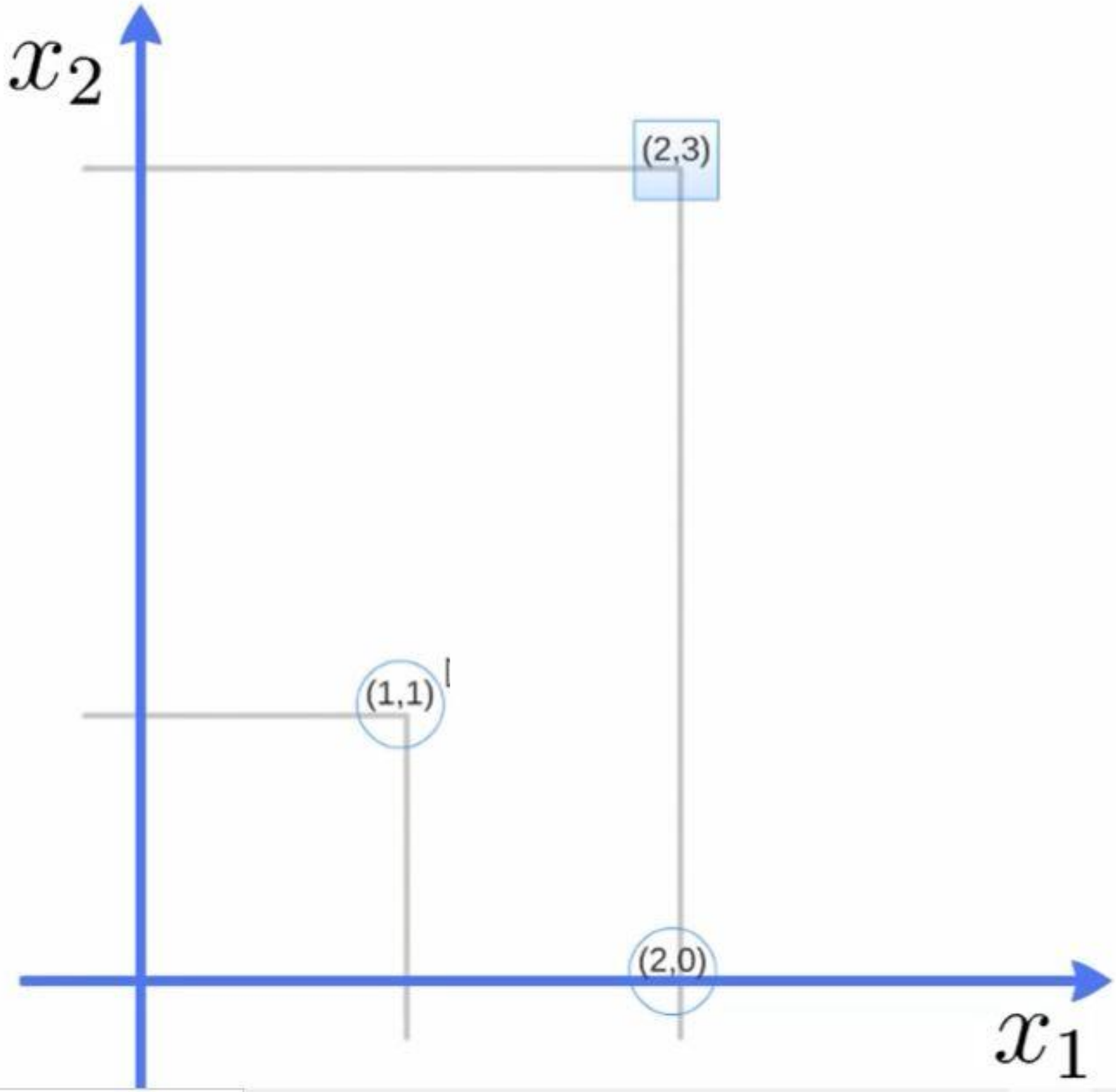


$$g(\vec{x}) \geqslant 1, \quad \forall \vec{x} \in \text{class } 1$$

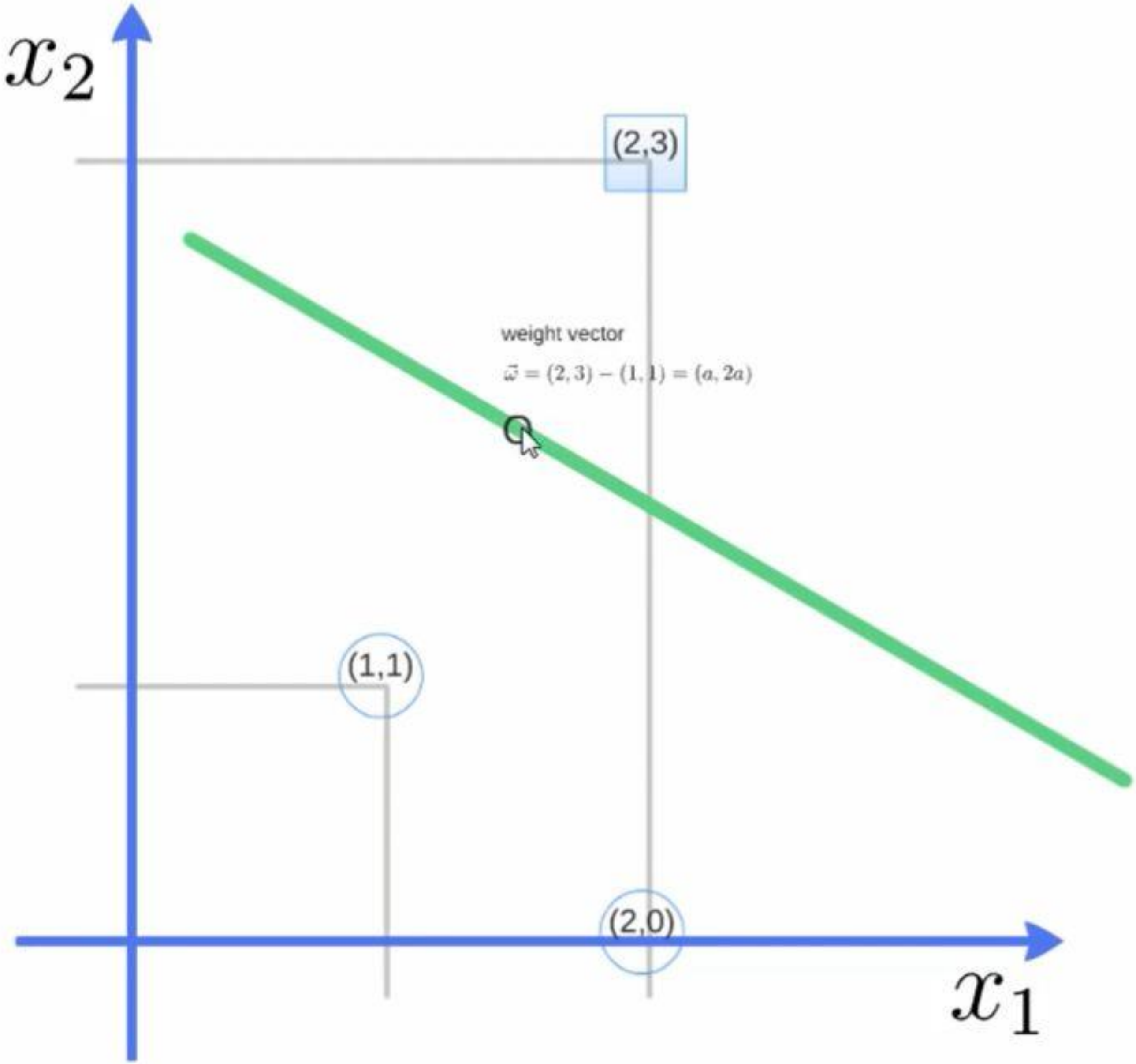
$$g(\vec{x}) \leqslant -1, \quad \forall \vec{x} \in \text{class } 2$$



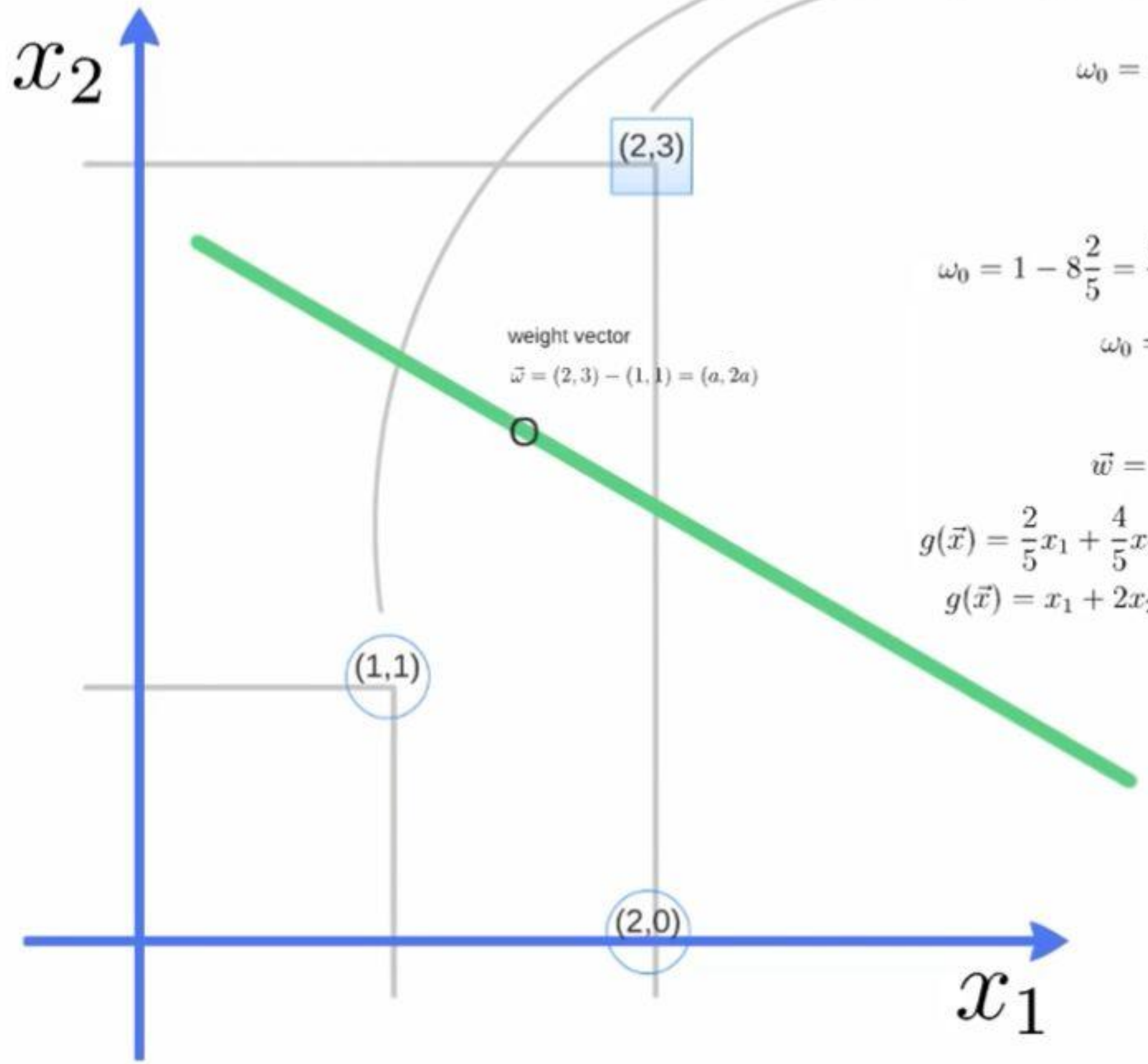
Example



Example



Example



weight vector $\vec{w} = (a, 2a)$

$a + 2a + \omega_0 = -1,$ using point (1,1)

$2a + 6a + \omega_0 = 1,$ using point (2,3)

$\dots$

$\omega_0 = 1 - 8a \qquad 3a + 1 - 8a = -1$

$\vdots \qquad 5a = 2$

$a = \frac{2}{5}$

$\omega_0 = 1 - 8\frac{2}{5} = \frac{5 - 16}{5}$

$\omega_0 = -\frac{11}{5}$

$\dots$

$\vec{w} = (\frac{2}{5}, \frac{4}{5})$

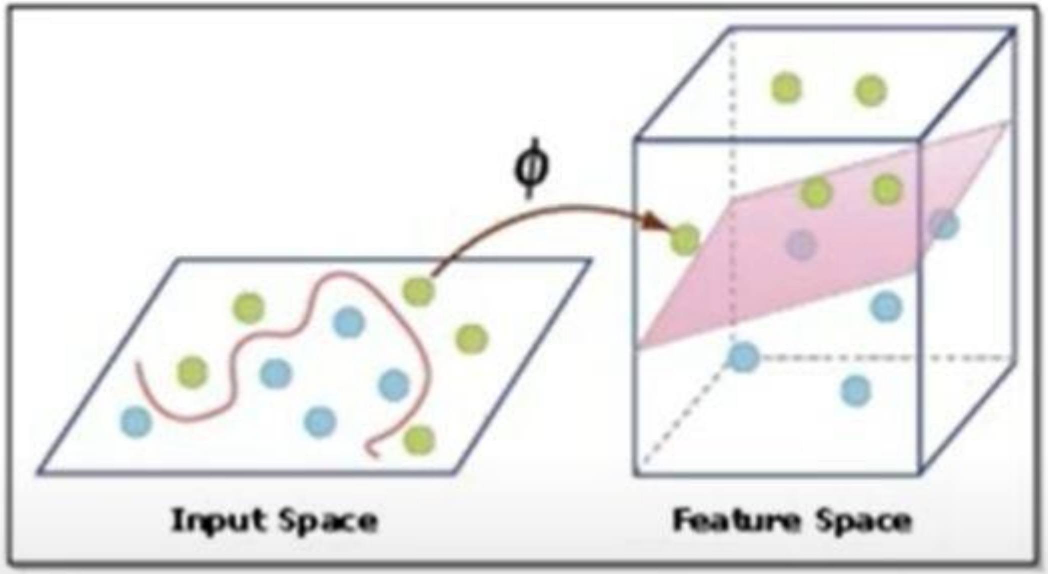
$g(\vec{x}) = \frac{2}{5}x_1 + \frac{4}{5}x_2 - \frac{11}{5}$

$g(\vec{x}) = x_1 + 2x_2 - 5.5$

$g(x_1,x_2) = a*x_1 + b*x_2 + w_o =$

$a*x_1 + 2a*x_2 + w_o$

Kernels

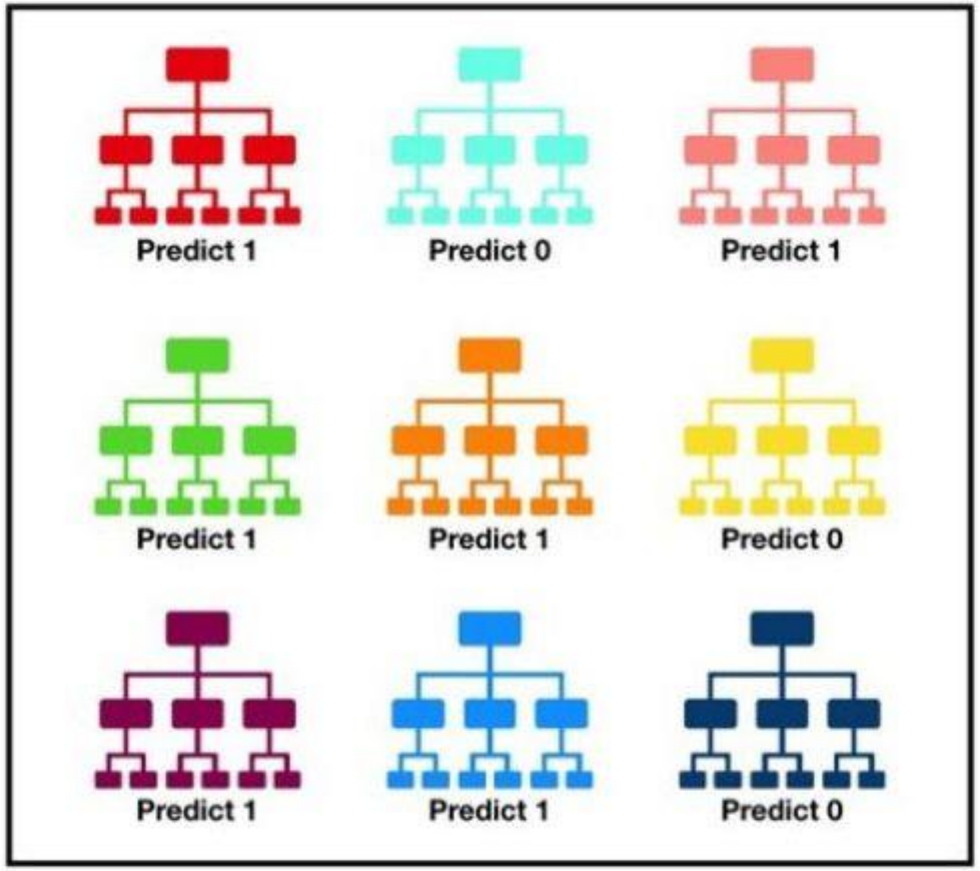


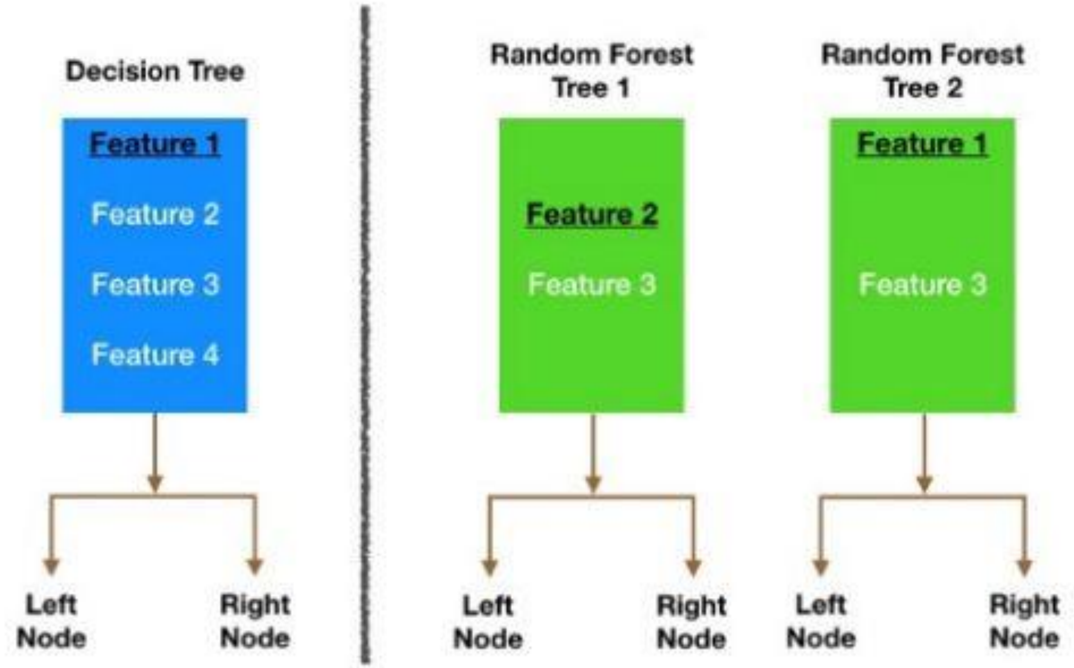
Kernel helps to find a hyperplane in the higher dimensional space without increasing the computational cost much

The Random Forest Classifier

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction (see figure below).

Bagging (Bootstrap Aggregation) — Decisions trees are very sensitive to the data they are trained on — small changes to the training set can result in significantly different tree structures. Random forest takes advantage of this by allowing each individual tree to randomly sample from the dataset with replacement, resulting in different trees. This process is known as bagging.





Node splitting in a random forest model is based on a random subset of features for each tree.

Feature Randomness — In a normal decision tree, when it is time to split a node, we consider every possible feature and pick the one that produces the most separation between the observations in the left node vs. those in the right node. In contrast, each tree in a random forest can pick only from a random subset of features. This forces even more variation amongst the trees in the model and ultimately results in lower correlation across trees and more diversification.