Advanced Machine Learning and Deep Learning

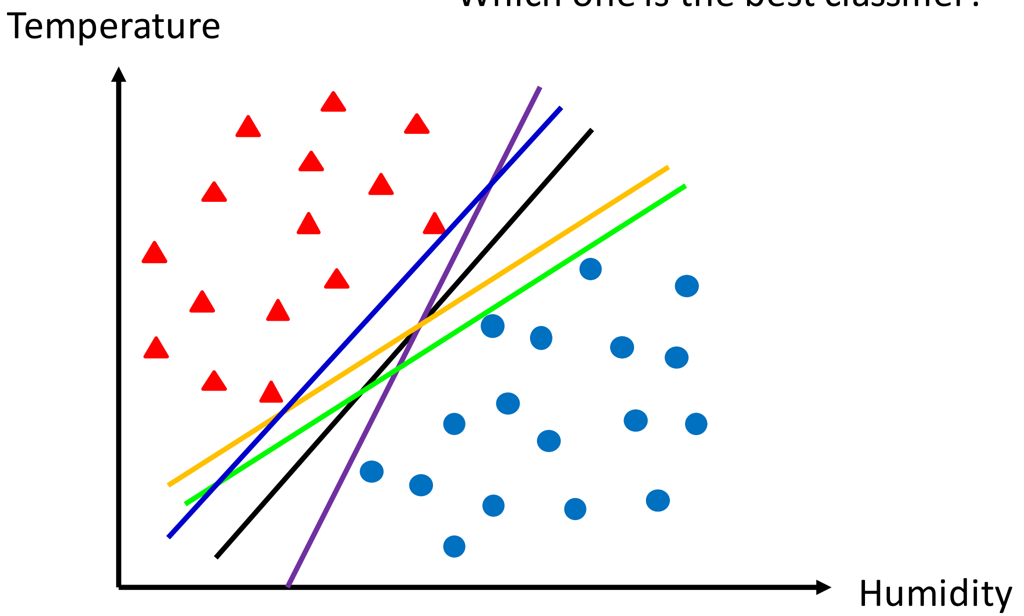
Dr. Mohammad Pourhomayoun
Assistant Professor
Computer Science Department
California State University, Los Angeles

Support Vector Machine (SVM)

Review: Support Vector Machine (SVM)

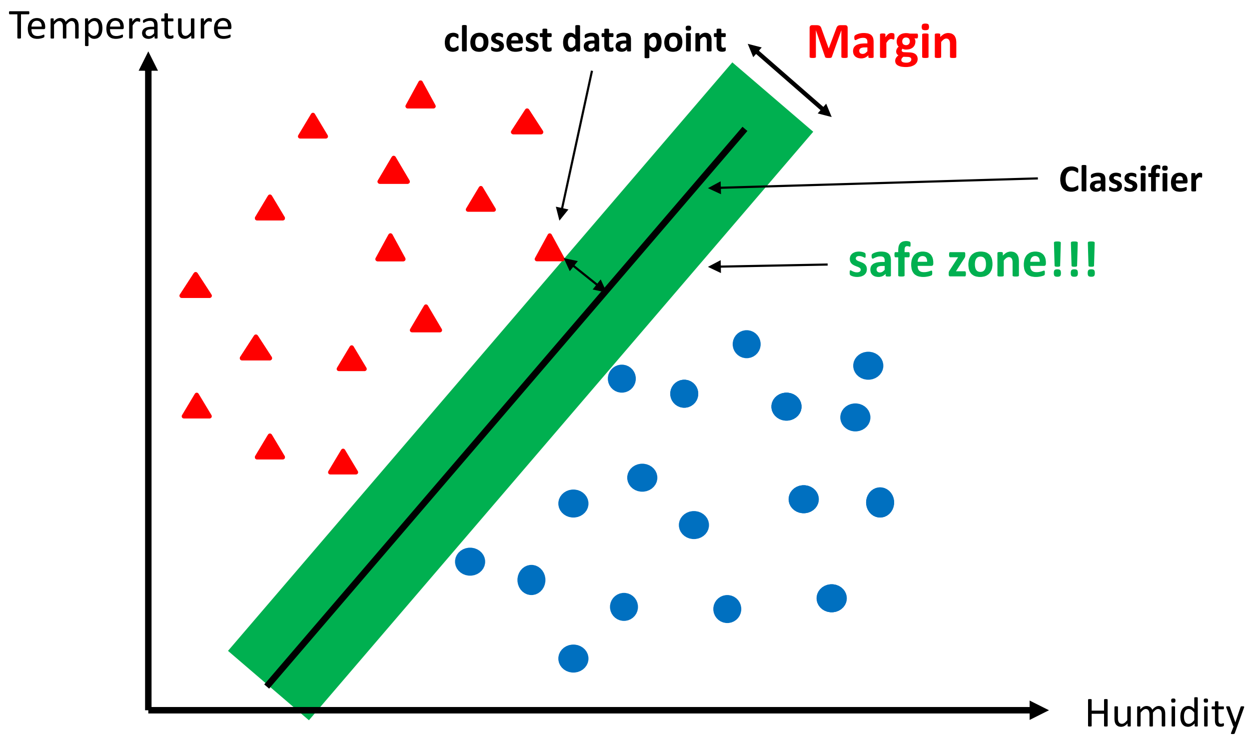
- SVM tries to constructs the best ***hyperplane***, or set of hyperplanes in feature space, which can be used for classifying data samples.
- Note: A 2D Hyperplane is a ***line*** in 2D space, A 3D Hyperplane is a ***plane*** in 3D space, … . In N-dimensional space, we just call it a Hyperplane.

- Which one is the best classifier?



Support Vector Machine (SVM)

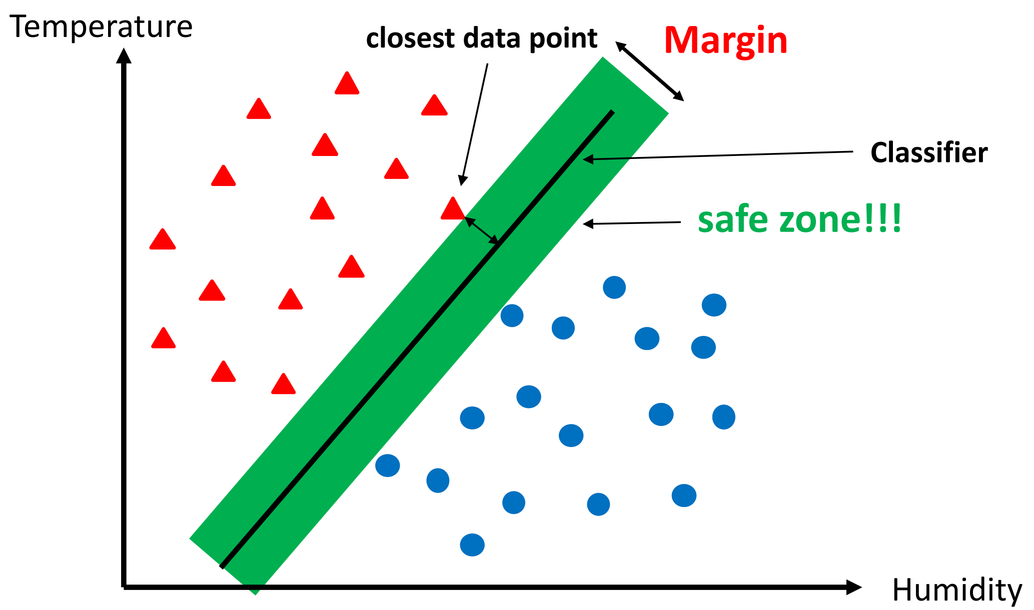
- Intuitively, a good separator (classifier) can be the hyperplane that has the ***largest separation***, or in other word, the largest distance to the nearest data samples of any class. This is called the ***largest margin (or maximum margin)***.



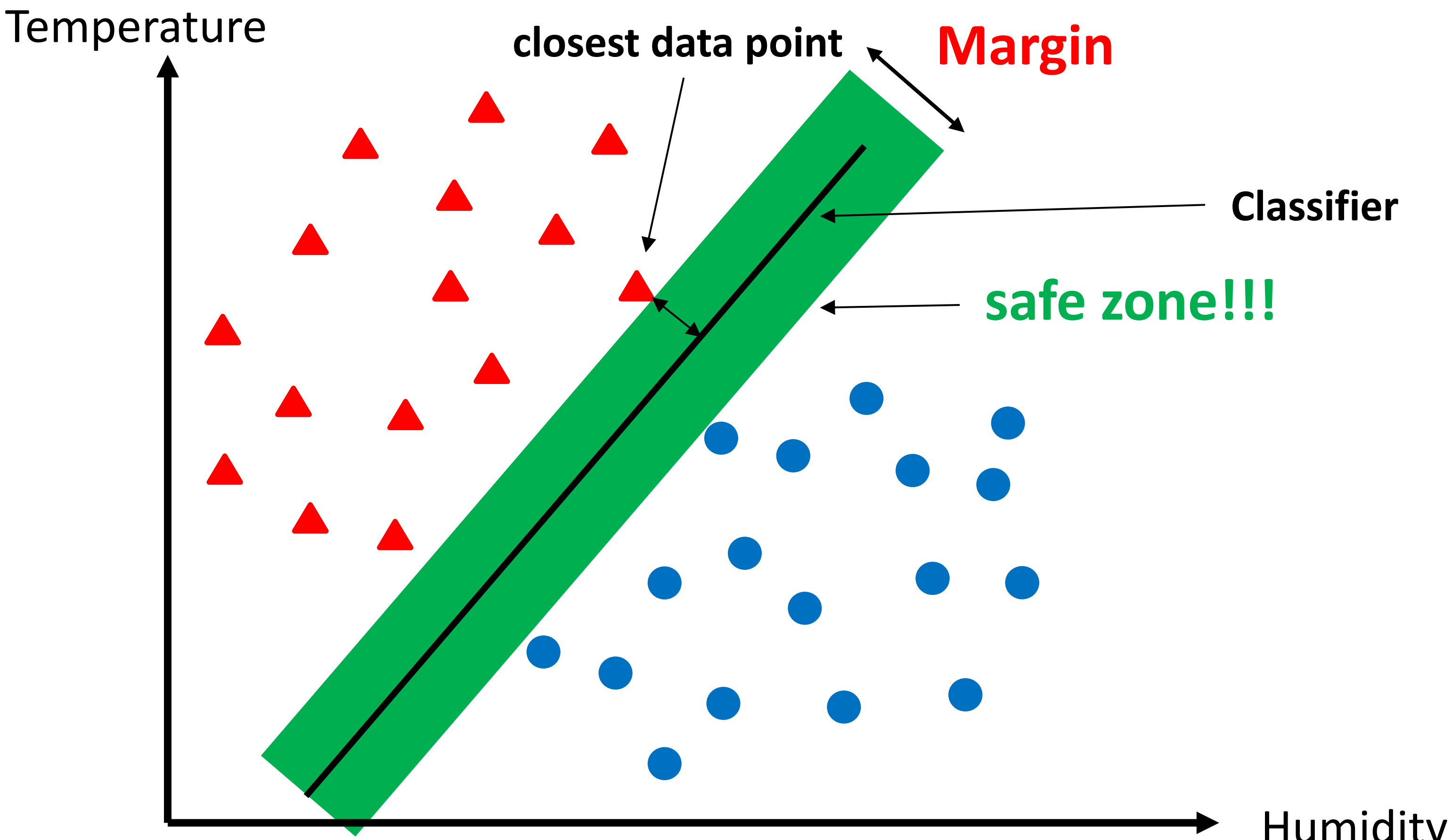
Support Vector Machine (SVM)

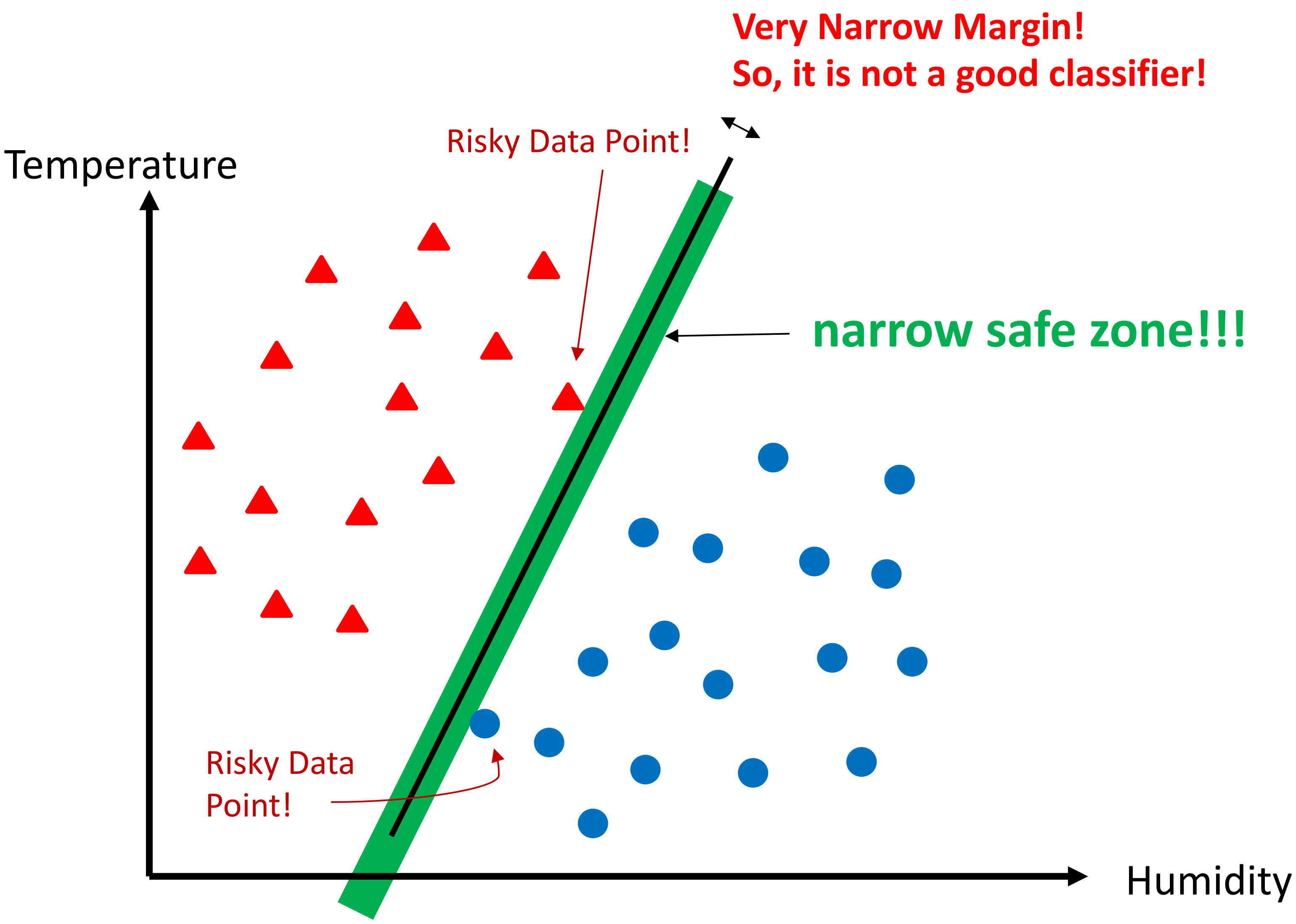
- Given a decision boundary (hyperplane), we can calculate the distance between the hyperplane and ***the closest data point***. If we double this distance, it is called the ***margin***. Thus, ***margin is the boundary area with no data point inside it!***

- **We would like to *maximize the margin* because the data samples near the decision boundary (near the hyperplane) represent very uncertain classification decisions (because any small noise or error can move them to the other side of the line!!!)**



- The best classifier is the one with the **Maximum Margin.**
- **Margin** is widest boundary area before hitting a data point.




Very Narrow Margin!
So, it is not a good classifier!
Risky Data Point!
Temperature
narrow safe zone!!!
Risky Data Point!
Humidity

Support Vector Machine (SVM)

- **Big Question:** How to find the hyperplane (the classifier) corresponding to the Maximum Margin?

- **<u>General equation for a 2-D line</u>:**

$$y = a\,x + b \qquad \text{or} \qquad w_1\,x_1 + w_2\,x_2 + b = 0$$

- **<u>General equation for an *m-dimensional linear hyperplane*</u>:**

$$w_1\,x_1 + w_2\,x_2 + \ldots + w_m\,x_m + b = 0$$

- **<u>Vector Representation</u>:**

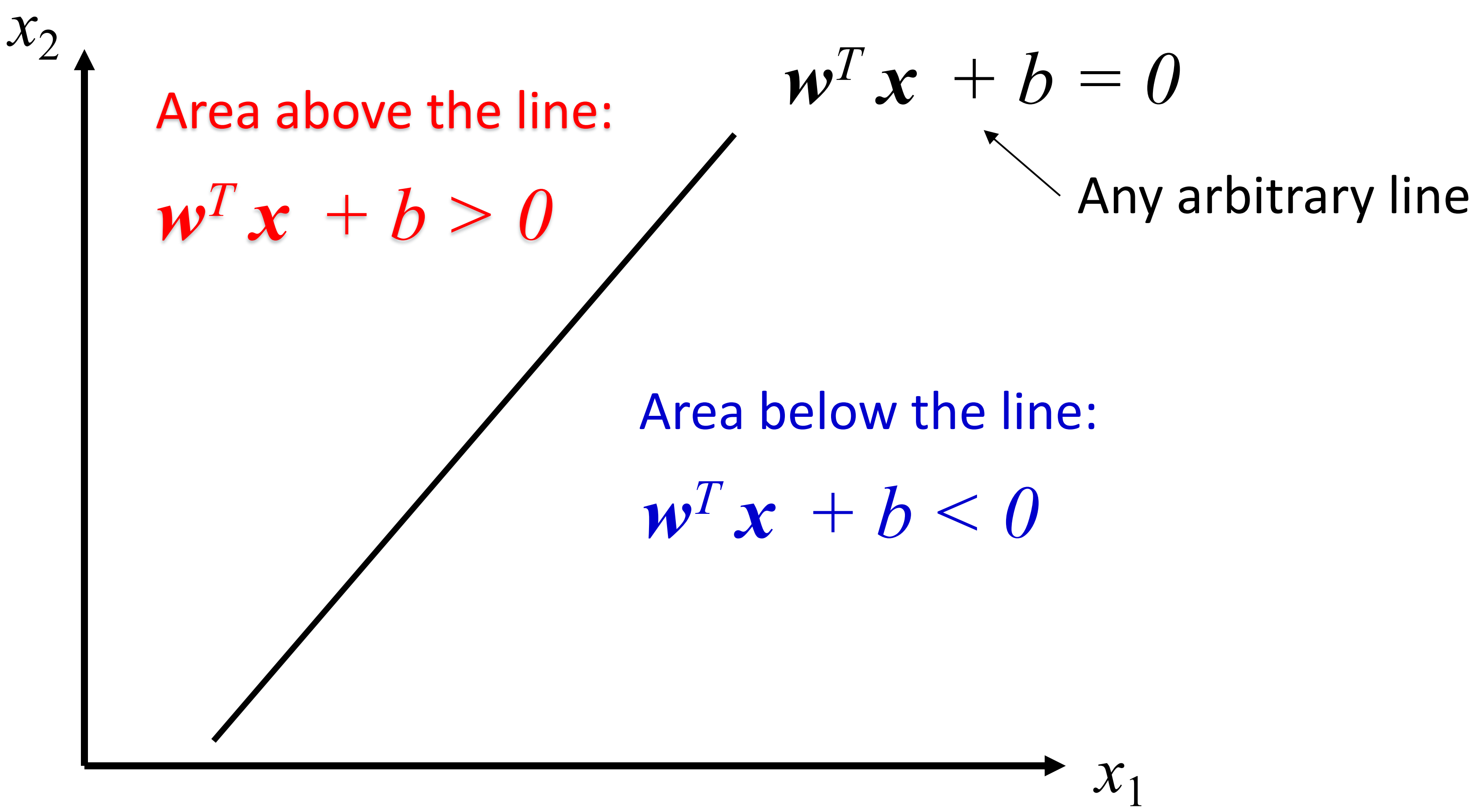
$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

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

Or equivalently:

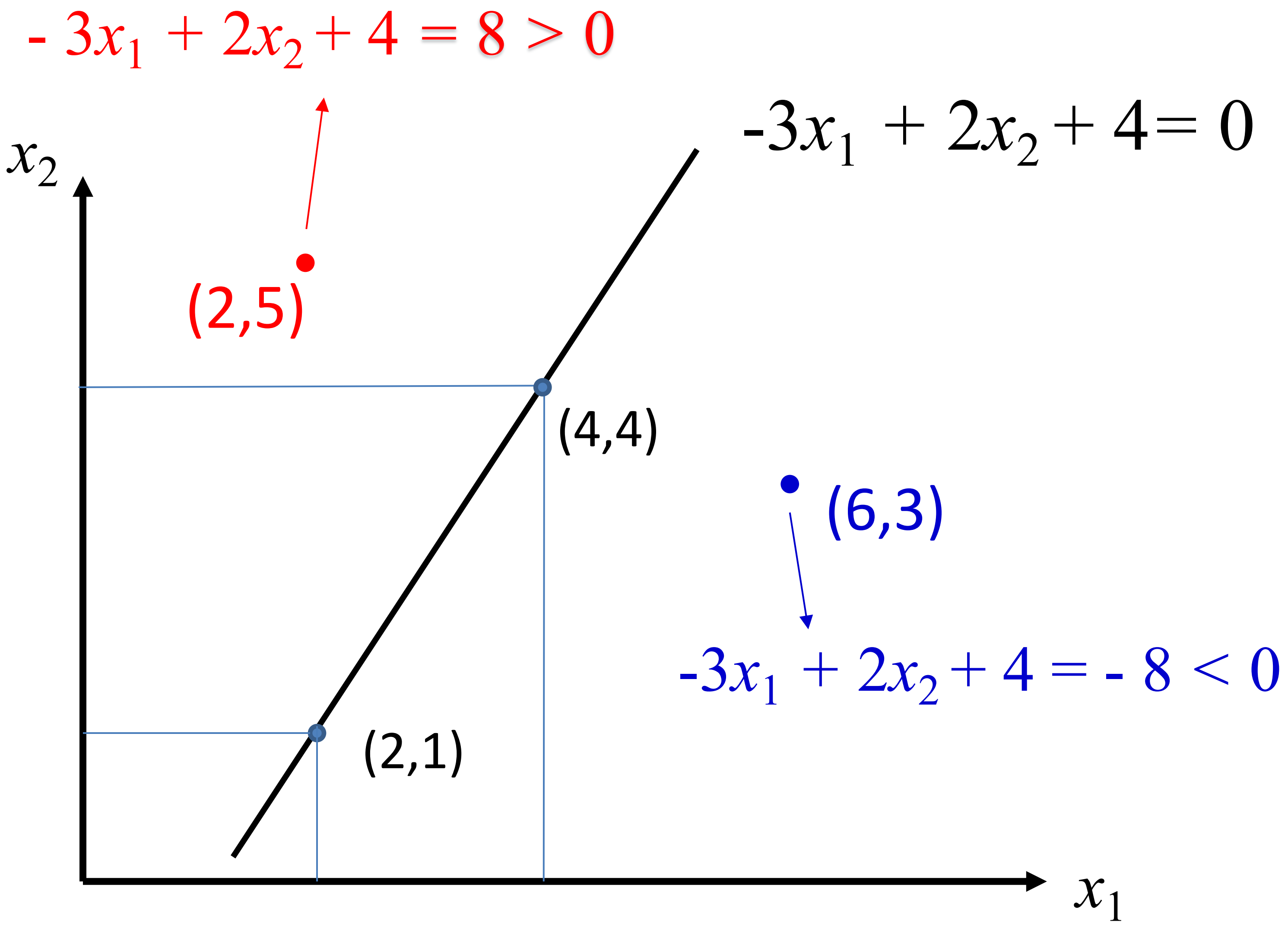
$$\boxed{\boldsymbol{w}.\boldsymbol{x} + b = 0}$$

Inner Product

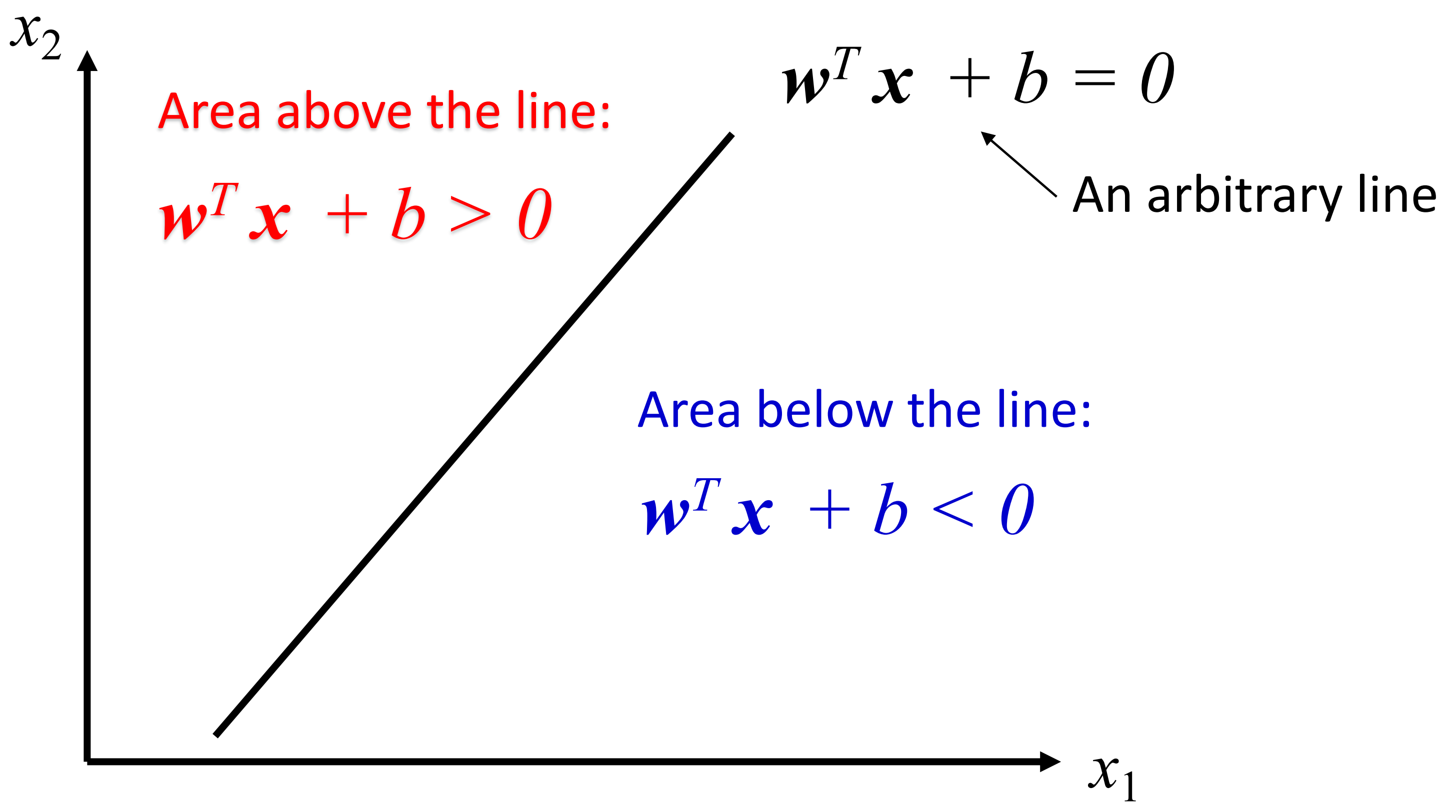


Note: The value of the equation has different signs in different sides of the line (left hand side is negative, right hand side is positive **OR VICE VERSA**)

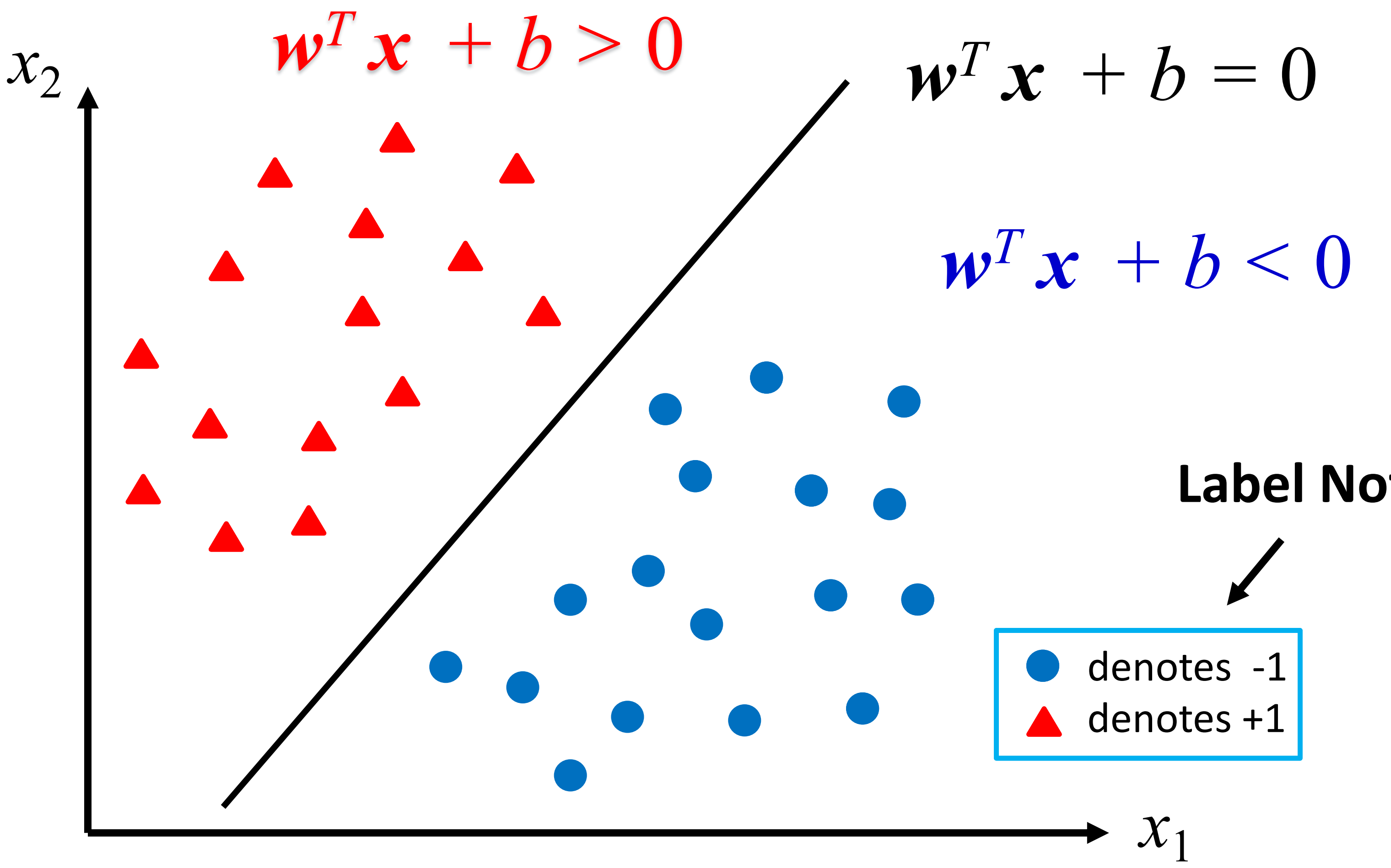
Example:



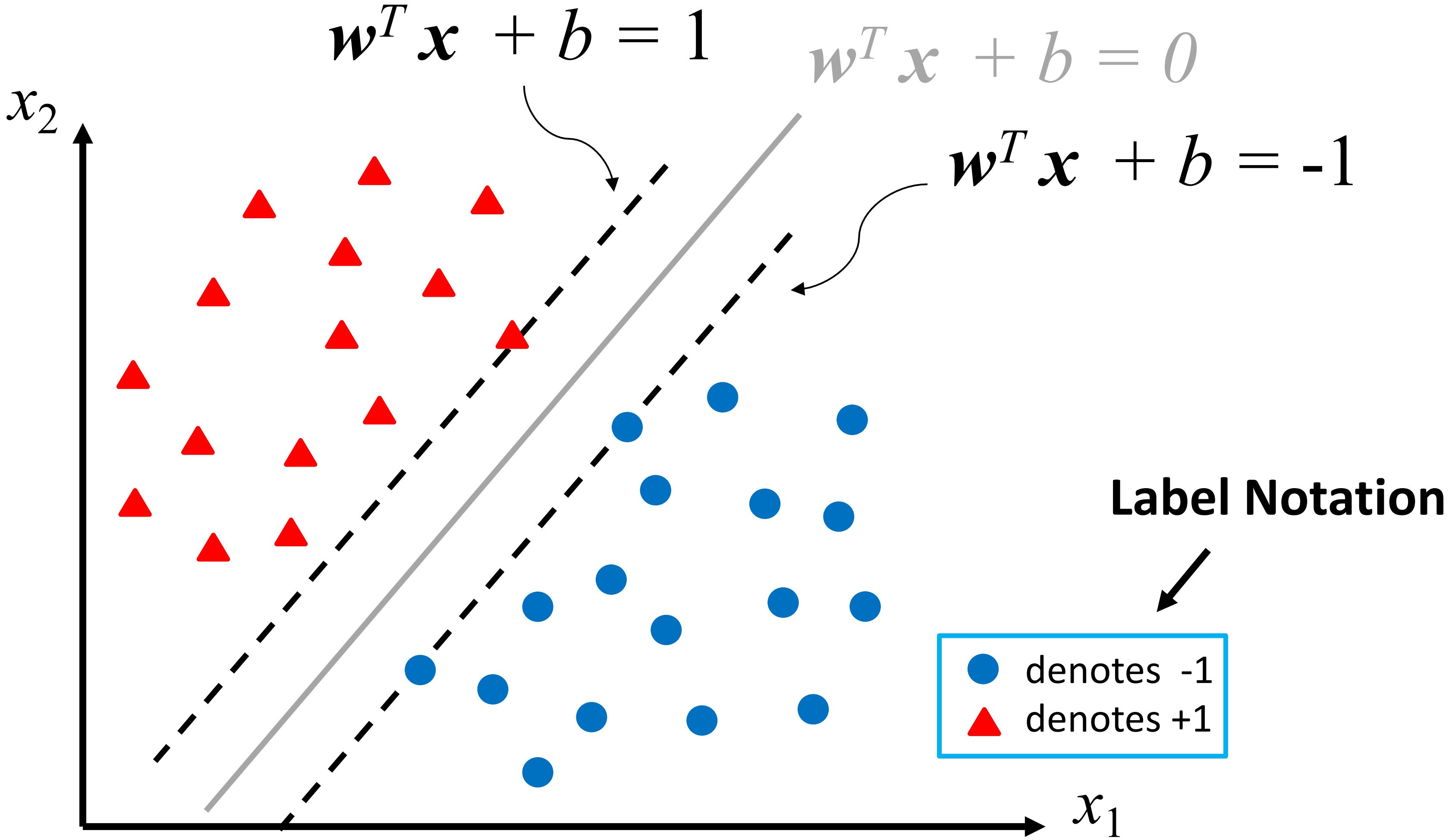
- **Question: How can we use this idea to build a classifier?**
    1. Training: Suppose that we have a set of training samples. How to find the classifier?
    2. Prediction: Suppose to have this classifier. How to make prediction for a new point?



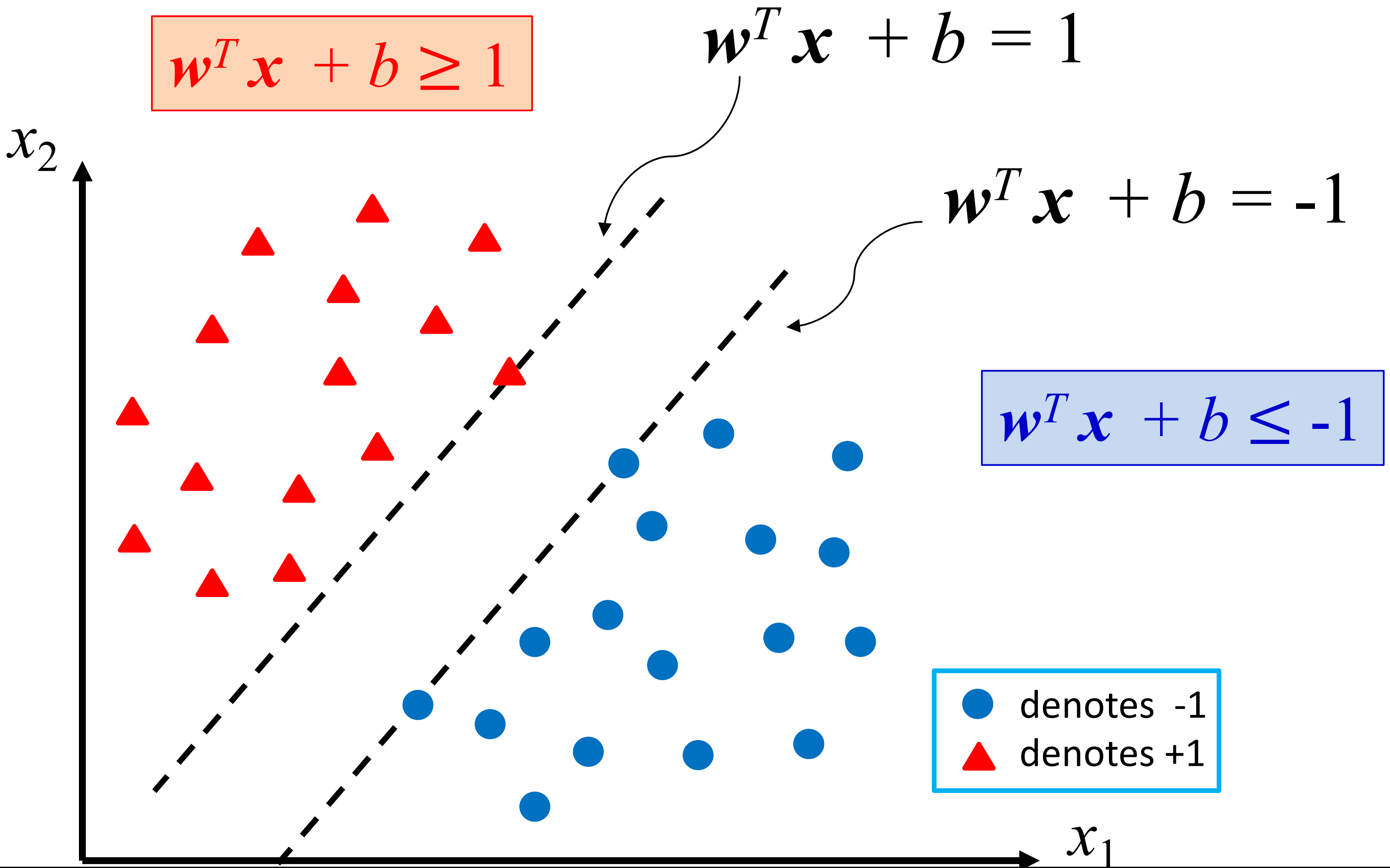
- Given a set of training data points, each point will be considered as a ***m-dimensional vector*** in space. We are looking for ***a line whose value is positive for red samples, and negative for blue samples.***

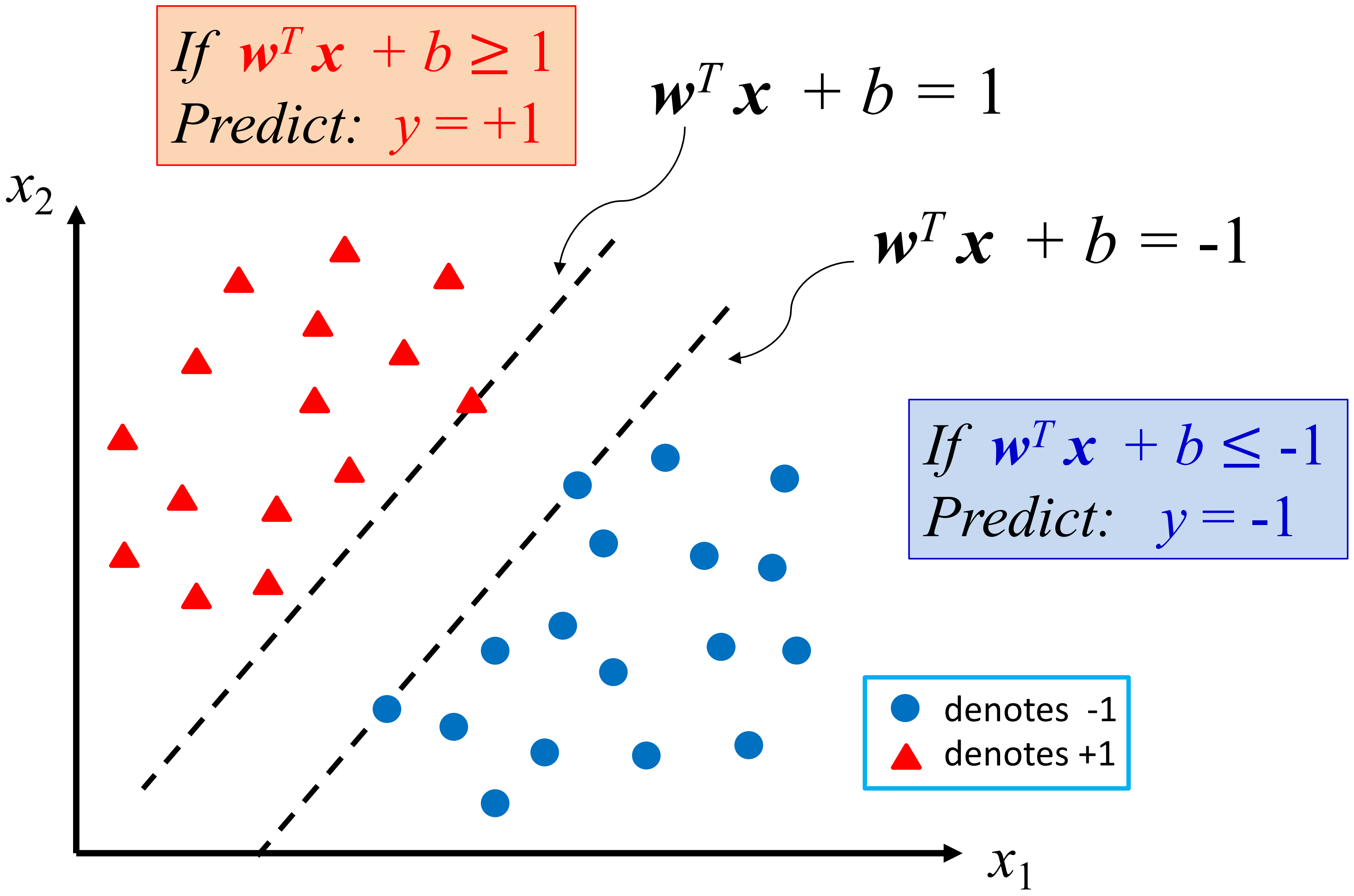


Let's do even better: We can find **two parallel lines** (rather than just one line) that separate the two classes of data! Red samples are ABOVE the TOP line, and blue samples are BELOW the BOTTOM line. These two hyperplanes can be described by the following equations:

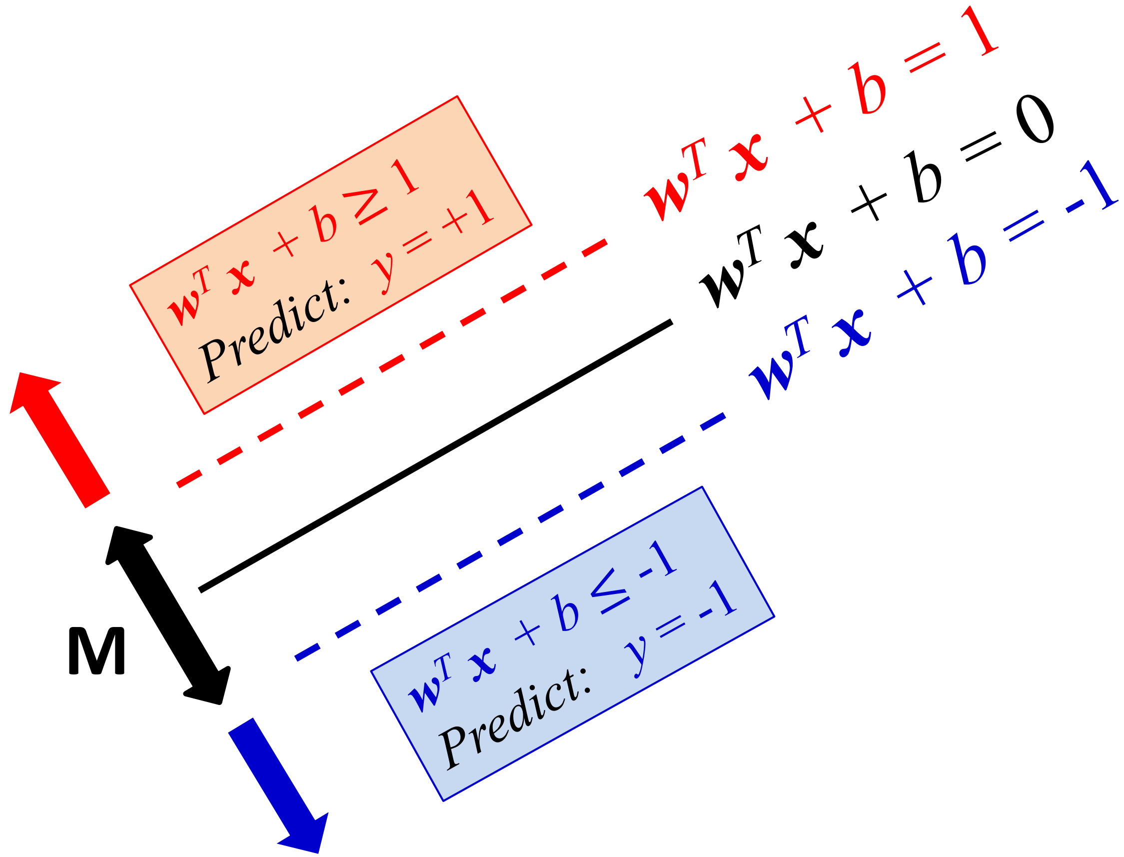


Now, Let's do even better than better: Now, let's find the **two parallel lines** that separate the two classes, **AND** the distance between them is as large as possible (**maximum margin**).





$If\ \boldsymbol{w}^T\boldsymbol{x} + b \geq 1$
$Predict:\ y = +1$
$\boldsymbol{w}^T\boldsymbol{x} + b = 1$
$\boldsymbol{w}^T\boldsymbol{x} + b = -1$
$If\ \boldsymbol{w}^T\boldsymbol{x} + b \leq -1$
$Predict:\ y = -1$
denotes -1
denotes +1
x_2
x_1



- M (the Margin) is the distance between the two lines "$\boldsymbol{w}^T \boldsymbol{x} + b = 1$" and "$\boldsymbol{w}^T \boldsymbol{x} + b = -1$".
- It is possible to show that the distance between these two line is $M = \frac{2}{||\boldsymbol{w}||}$.

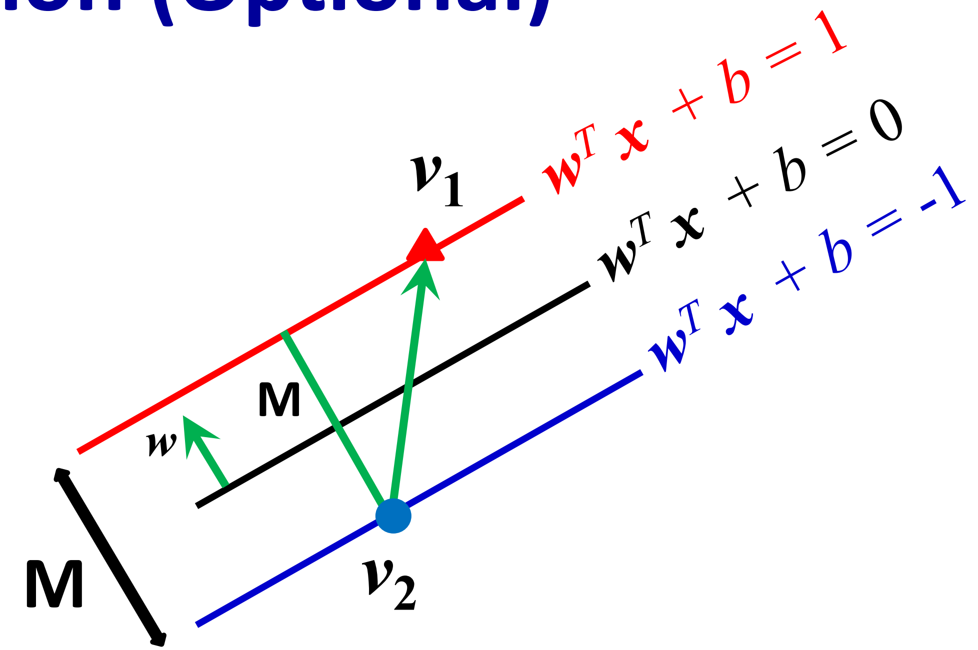
Margin Calculation (Optional)

Projection:

$$M = \| \boldsymbol{v}_1 - \boldsymbol{v}_2 \| \, \mathrm{Cos}\,\theta$$

$$= \| \boldsymbol{v}_1 - \boldsymbol{v}_2 \| \frac{(\boldsymbol{v}_1 - \boldsymbol{v}_2).\boldsymbol{w}}{\| \boldsymbol{w} \| . \| \boldsymbol{v}_1 - \boldsymbol{v}_2 \|}$$

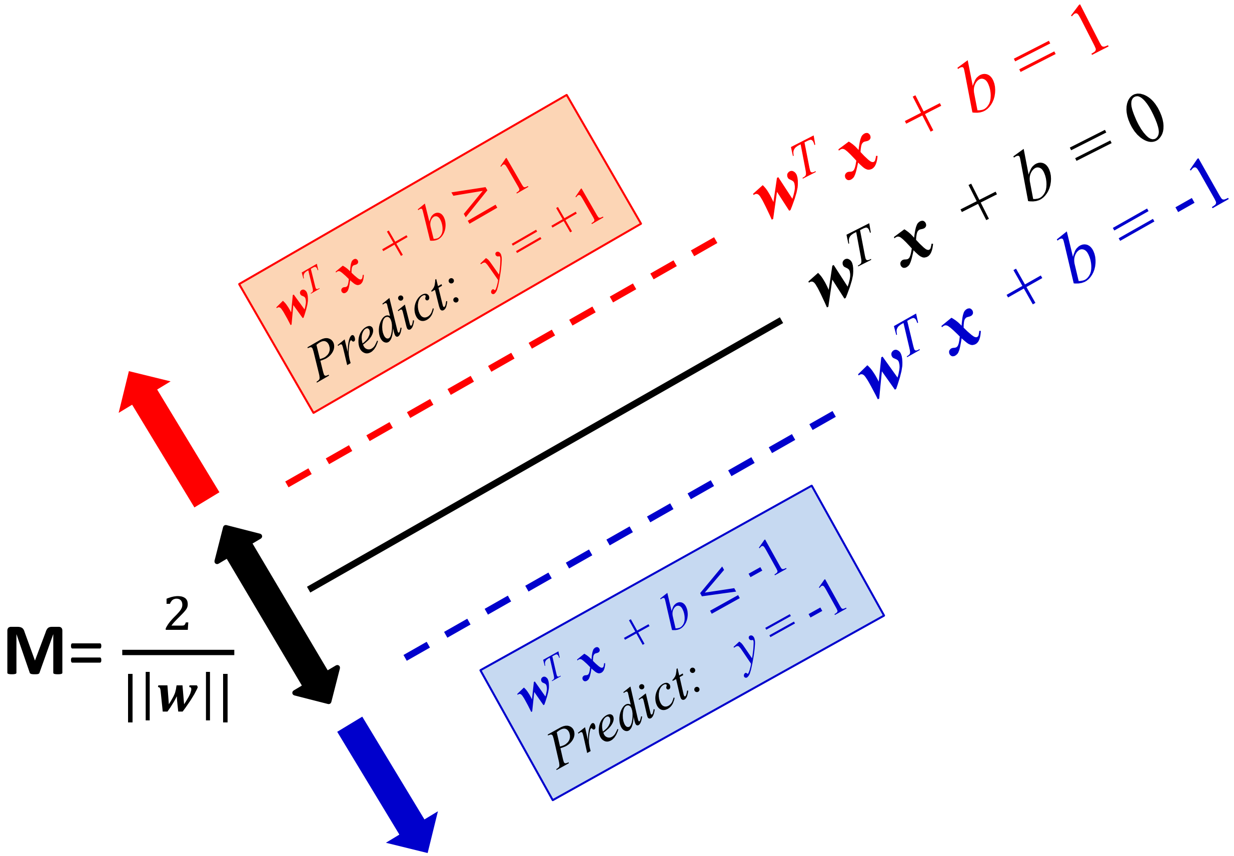
$$= (\boldsymbol{v}_1 - \boldsymbol{v}_2).\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \qquad (1)$$



$$\left.\begin{aligned} \boldsymbol{w}.\boldsymbol{v}_1 + b &= 1 \\ \boldsymbol{w}.\boldsymbol{v}_2 + b &= -1 \end{aligned}\right\} (\boldsymbol{v}_1 - \boldsymbol{v}_2).\ \boldsymbol{w} = 2 \quad (2)$$

Subtract

(1) and (2) $\rightarrow$ $M = \dfrac{2}{\|\boldsymbol{w}\|}$

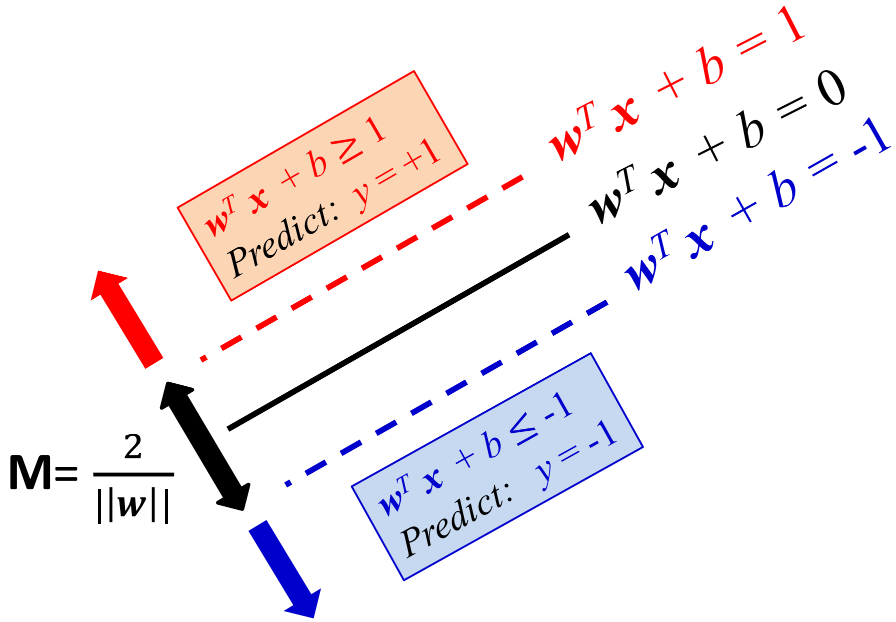


- Given a Training set, the goal is to find the best classifier that classifies the samples correctly and has the maximum margin (maximum M = $\frac{2}{||\boldsymbol{w}||}$).

SVM Classifier

- ***Formulation:***

$$\begin{cases} \text{Maximize } \dfrac{2}{\|\boldsymbol{w}\|} \\ \text{Such that:} \\ \text{for } y=+1, \ \boldsymbol{w}^T\boldsymbol{x} + b \geq 1 \\ \text{for } y= -1, \ \boldsymbol{w}^T\boldsymbol{x} + b \leq -1 \end{cases}$$



SVM Classifier

- ***Let's simplify this formulation a little bit:***

1- Maximize $M = \dfrac{2}{\|\boldsymbol{w}\|}$ is equivalent to Minimize $\dfrac{1}{2}\|\boldsymbol{w}\|^2$

2-

$$\left.\begin{array}{l} \text{for } y = +1,\ \boldsymbol{w}^T\boldsymbol{x} + b \geq 1 \\ \text{for } y = -1,\ \boldsymbol{w}^T\boldsymbol{x} + b \leq -1 \end{array}\right\}$$ is equivalent to: $\boxed{y\,(\boldsymbol{w}^T\boldsymbol{x} + b) \geq 1}$

The Optimization Problem

- Thus, given a training set $\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\}$, the new formulation is:

$$\begin{cases} \text{Minimize } \dfrac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{Subject to: } y_i\,(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1 \\ \text{for } i = 1, 2, \ldots, n \end{cases}$$

Thank You!

Questions?