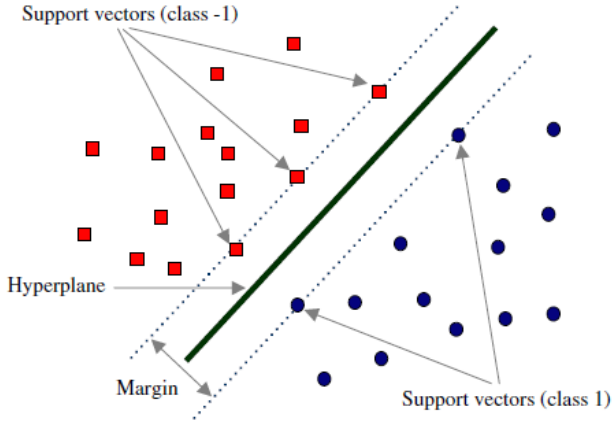
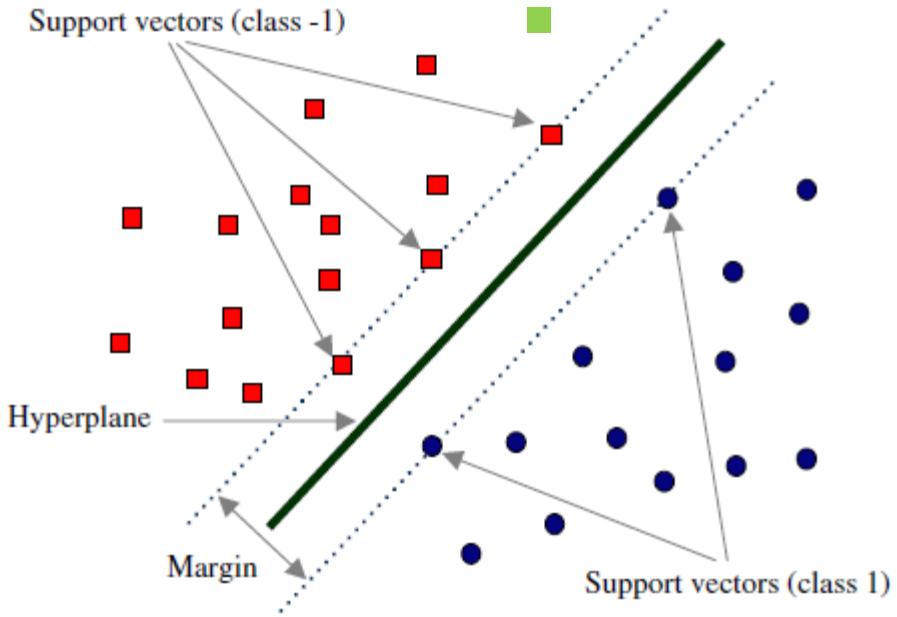
Siddhardhan

Math behind Support Vector Machine (SVM) Classifier

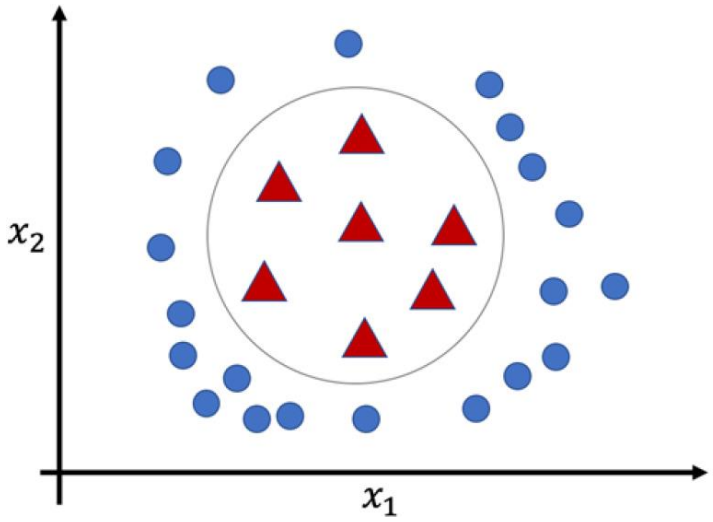


Support Vector Machine Classifier



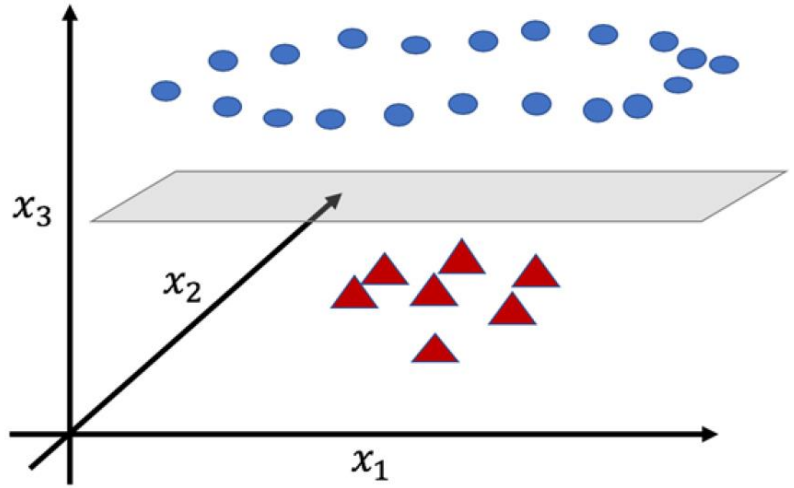
- Hyperplane
- Support Vectors
- Margin
- Linearly separable data

Support Vector Machine Classifier



SVM in 2 dimensions

Kernel



SVM in 3 dimensions

Support Vector Machine Classifier

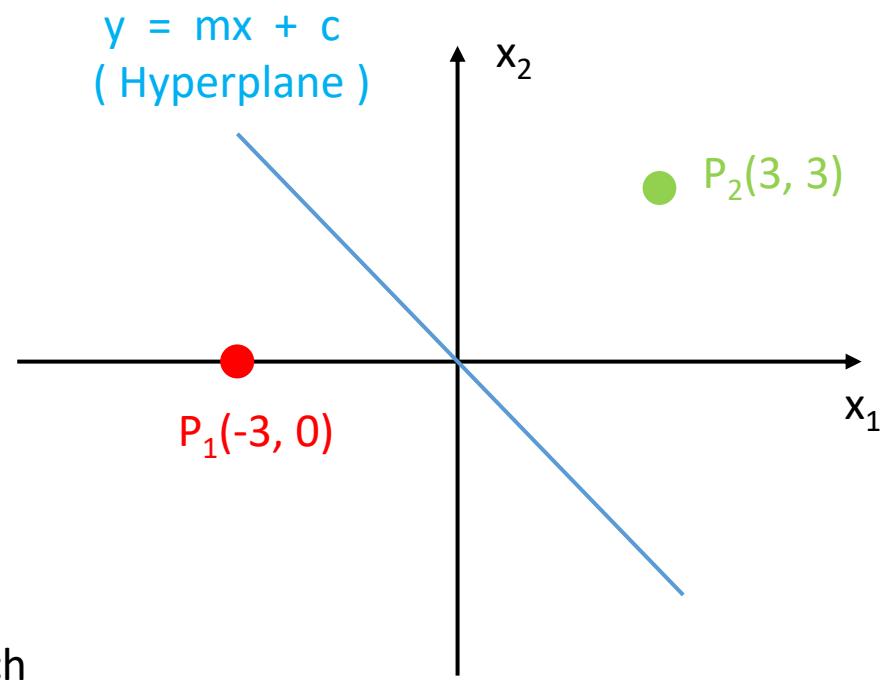
$P_1(-3, 0)$

$w^T x = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -3 & 0 \end{bmatrix}$

$w^T x = 3$

(Positive)

Inference: For all the points which lie in the left side of the hyperplane, $w^T x$ value will be Positive



Let slope, m = - 1

Intercept, c = 0

w --> parameters of the line

(m, c) = (-1, 0)

$P_2(3, 3)$

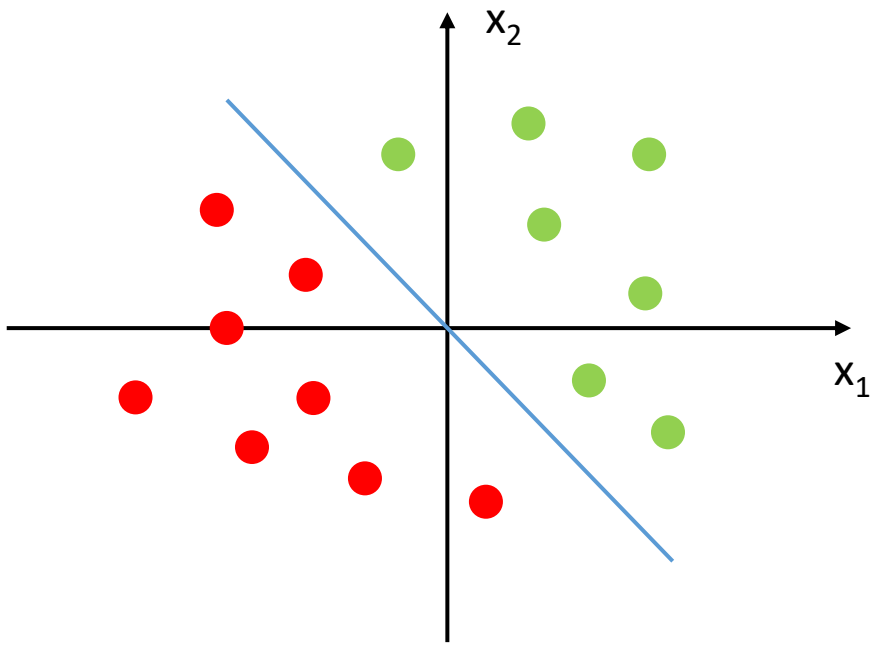
$w^T x = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 3 & 3 \end{bmatrix}$

$w^T x = -3$

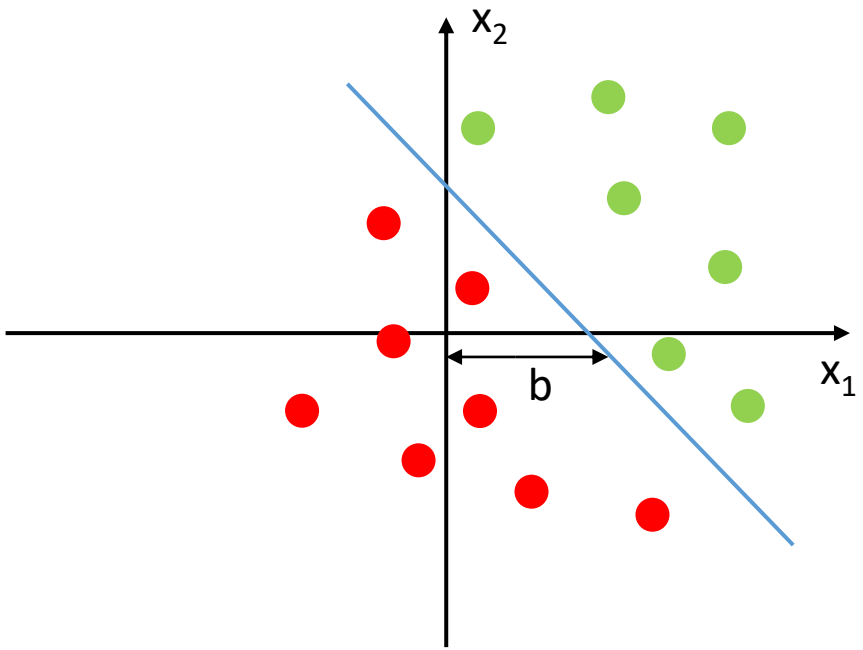
(Negative)

Inference: For all the points which lie in the right side of the hyperplane, $w^T x$ value will be Negative

Support Vector Machine Classifier



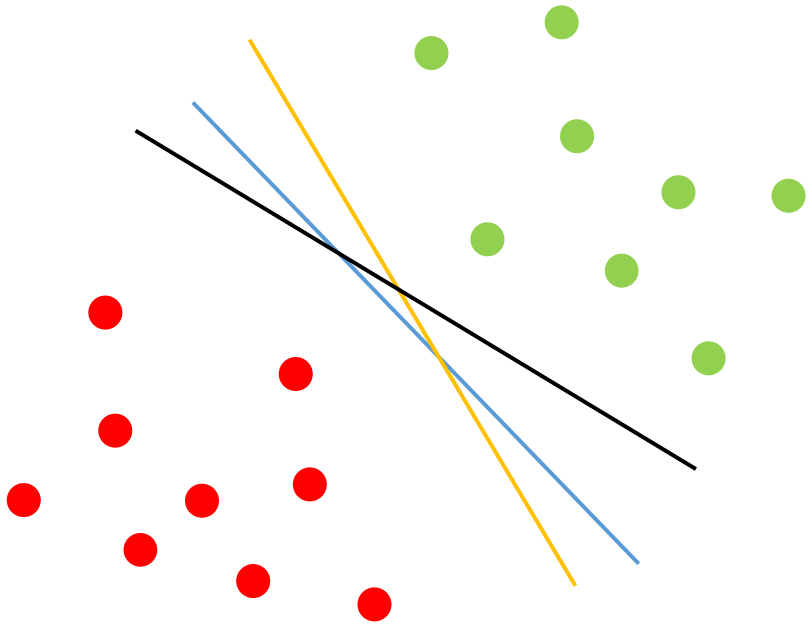
$w^T x = \text{Label}$



$w^T x + b = \text{Label}$

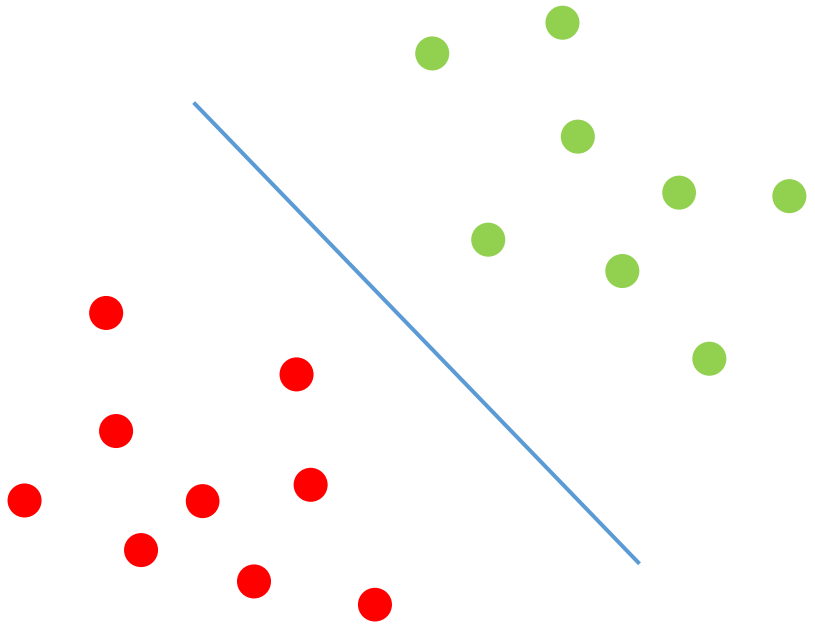
Support Vector Machine Classifier

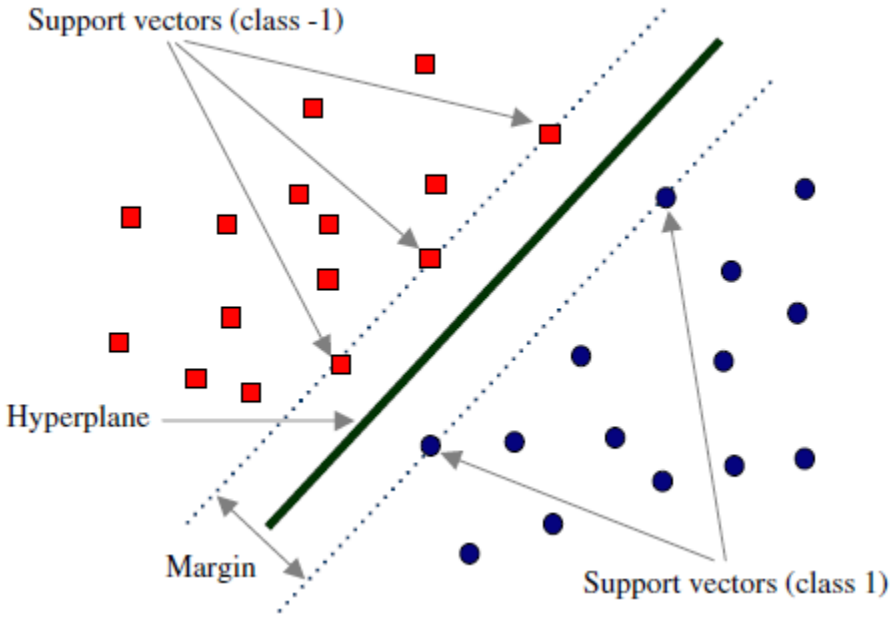
Which is the best Hyperplane?



Support Vector Machine Classifier

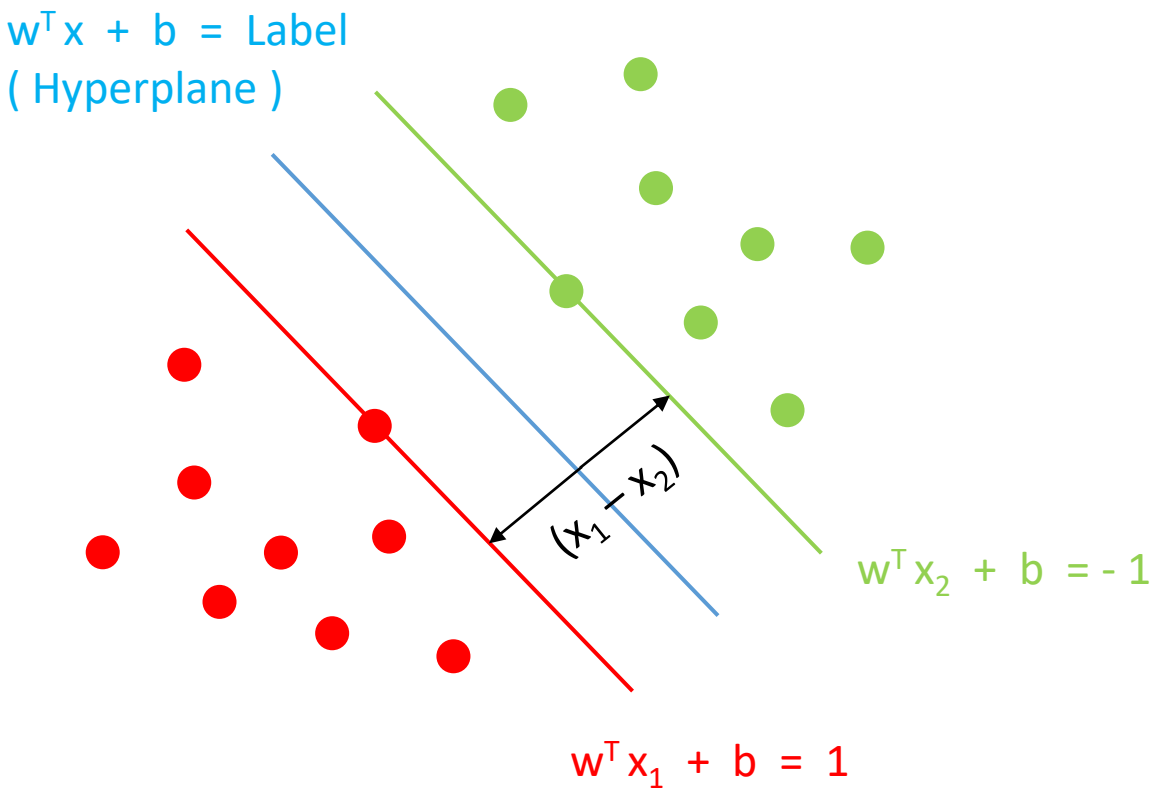
Which is the best Hyperplane?





Support Vector Machine Classifier

Optimization for Maximum margin:



$$w^T x_1 + b = 1$$

$$(-)\quad w^T x_2 + b = -1$$

$$w^T (x_1 - x_2) = 2$$

Divide by $||w||$

(magnitude of the vector)

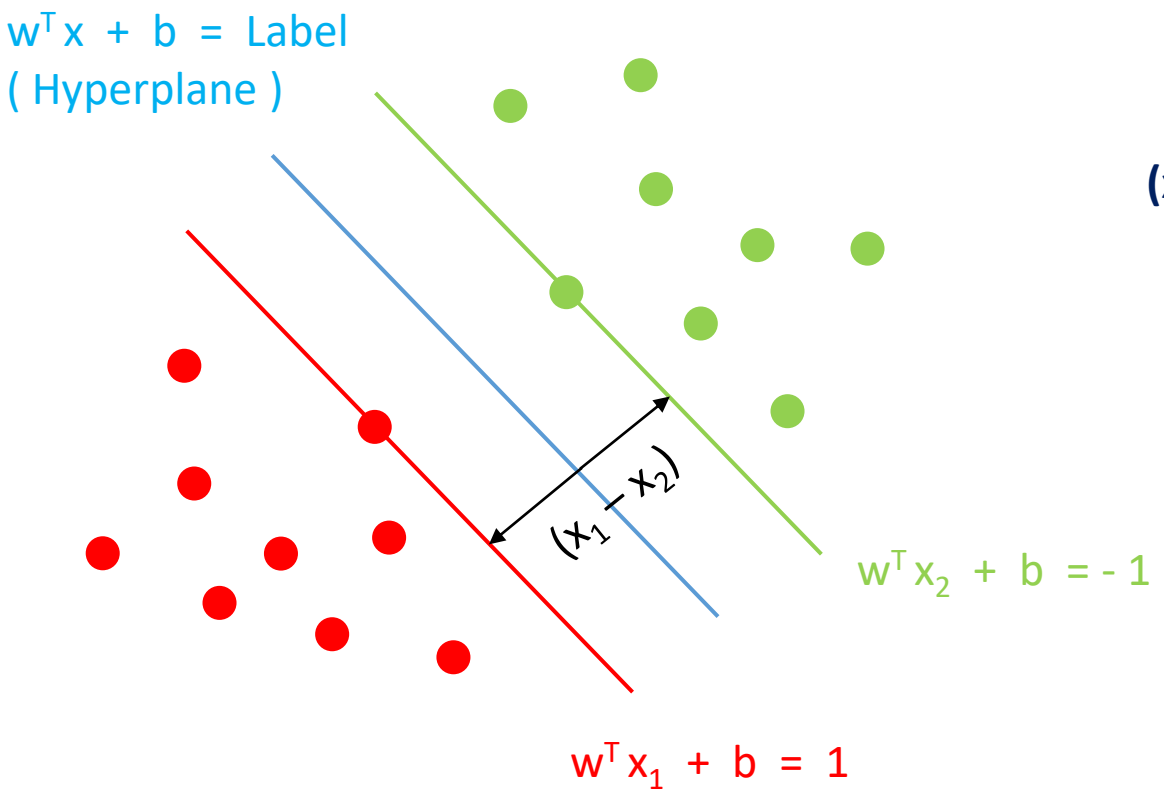
$$\frac{w^T}{||w||}(x_1 - x_2) = \frac{2}{||w||}$$

$$\mathbf{(x_1 - x_2) = \frac{2}{||w||}}$$ (margin)

Support Vector Machine Classifier

Optimization for Maximum margin:

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases} \quad \text{(Label)}$$



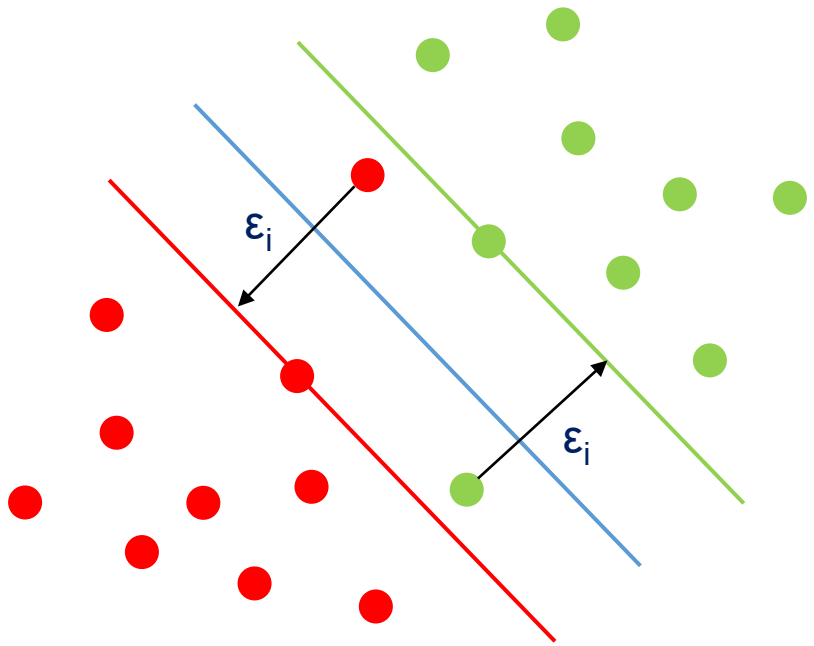
$$(x_1 - x_2) = \frac{2}{||w||} \quad \text{(margin)}$$

$$\max \left(\frac{2}{||w||} \right) \quad \text{Such that,}$$

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases}$$

Support Vector Machine Classifier

Maximum margin without overfitting:



$$\max \left(\frac{2}{||w||} \right) \quad \text{Such that,}$$

$$y_i = \begin{cases} -1, & w^T x_1 + b \leq -1 \\ 1, & w^T x_1 + b \geq 1 \end{cases}$$

$$\min \left(\frac{||w||}{2} \right) + c * \Sigma \varepsilon_i$$

c --> Number of errors

ε_i --> Error magnitude