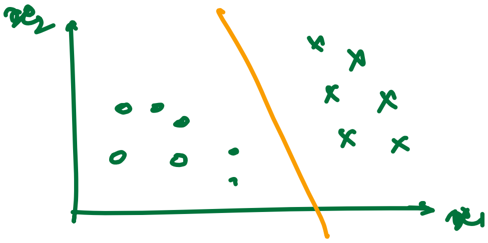
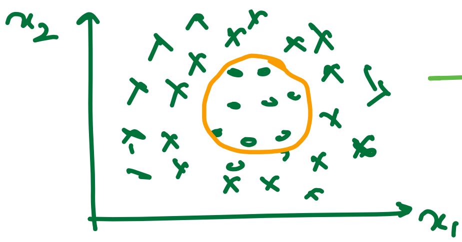
SVM (Support Vector Machine)

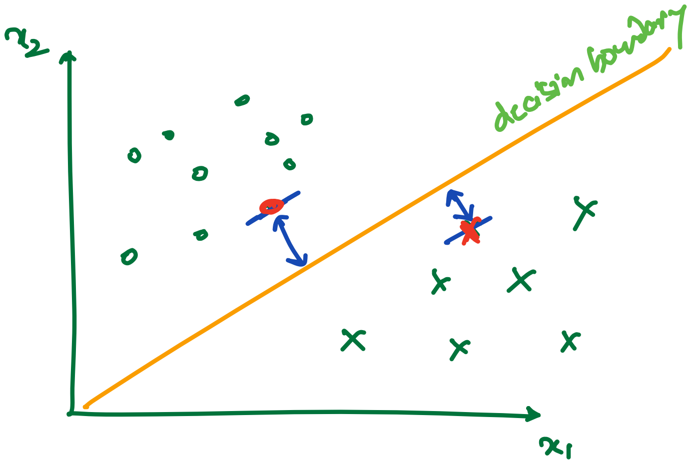
↳ classifier → linearly separable data
→ non linearly separable data





→ Kernel trick

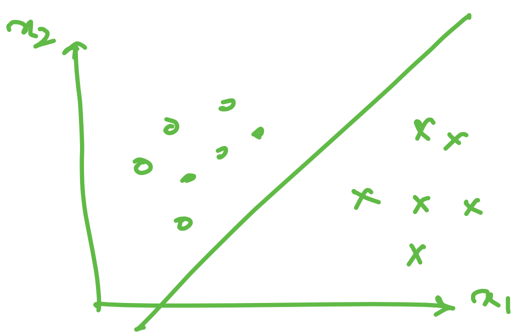
Logistic Regression



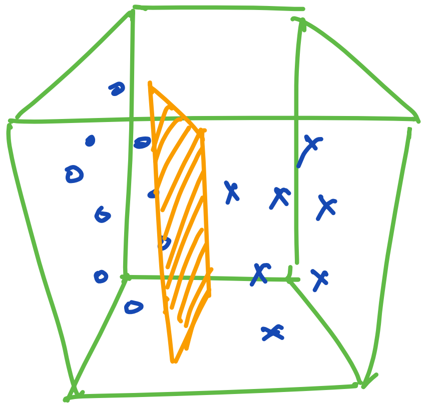
SVM:
Points which are closest to your decision boundary should be very far away from each other.

Hyperplane:

n features hyperplane will be of n-1 dimensions & it should separate both the classes.



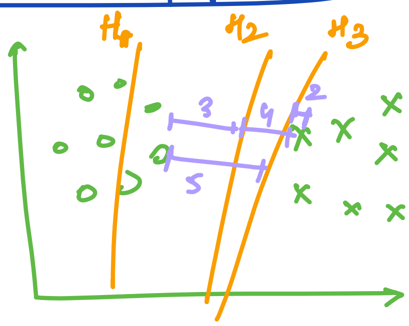
2 features:
db: line (1D)



3D features

2D hyperplane : plane

Separating Hyperplane:



$H_1 \rightarrow$ separating hyperplane ✗

H_2, H_3 ✓

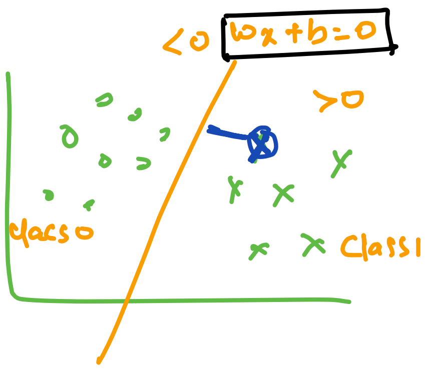
H_3 : (2, 5) — 2

H_2 : (3, 4) — 3 ✓

maximise it

minimum distance

hyperplane is a separating hyperplane if :



$y = mx + c$

$wx + b$

$\frac{wx+b}{\|w\|_2} > 0$

$wx+b > 0$ If data points are from class 1

$wx+b < 0$ if data points are from class 0

Linear Regression:

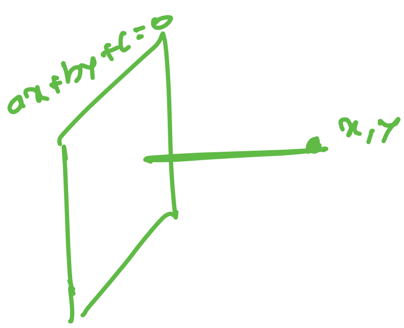
$\theta^T x = 0$ $\qquad mx + c = 0$

$x_1 \quad x_2 \quad x_3 \dots x_n$

$\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots$

$\theta_n x_n + \theta_0 x_0$

$x_0 \rightarrow 1$



$\frac{ax+by+c}{\sqrt{a^2+b^2}}$ normalised distance

$\rightarrow L_2$ norm

L_1 norm: $(a^1+b^1)^{1/1}$

L_3 norm:

$$(a^3+b^3)^{1/3}$$

L_2 norm:

$$(a^2+b^2)^{1/2}$$

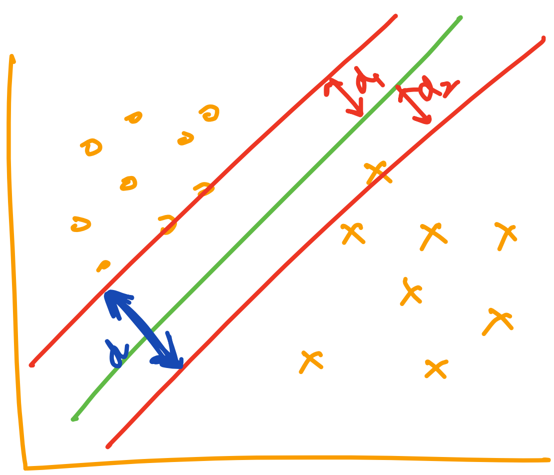
$wx+b \longrightarrow \underbrace{w_1x_1 + w_2x_2 + b}_{\text{bias}} = 0$

$x_1 x_2$ features
$w_1 w_2$ weights

$x_1^{(i)}$ $x_2^{(i)}$

$$\frac{w_1 x_1^{(i)} + w_2 x_2^{(i)} + b}{\sqrt{w_1^2 + w_2^2}}$$

$\|w\|_2$
L_2 norm



hyperplane with maximum margin

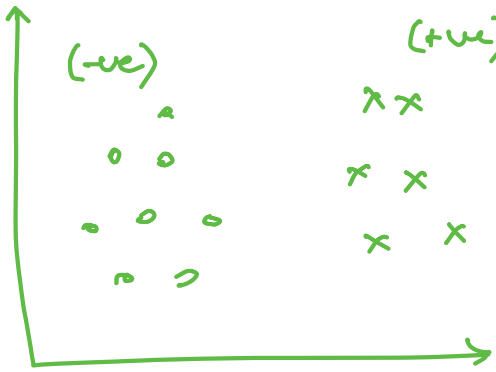
formulating objective

$X = \{x^1, x^2 \; ..... \; x^m\}$

$Y = \{y^1, y^2 \dots\dots y^m\}$

Binary Classification: $y^{(i)} \in \{-1, 1\}$

↳ class label will be 1 or -1

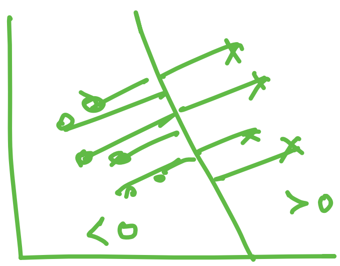


find a separating hyperplane with maximum margin.

$w^T x + b = 0$ → Hyperplane equation

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

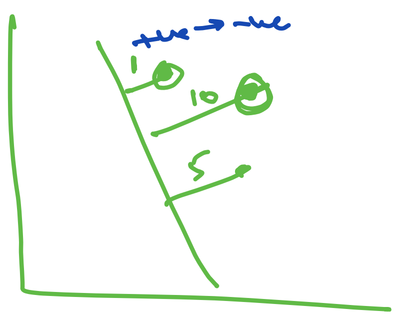
→ Separating hyperplane:



$w^T x + b > 0$ if $x \in$ +ve class

$w^T x + b < 0$ if $x \in$ -ve class

→ Confident about the predictions



↳ if you slightly change the line the distance would become -ve from +ve.

$$z = w^T x + b$$

$$y_{pred} = g(z)$$

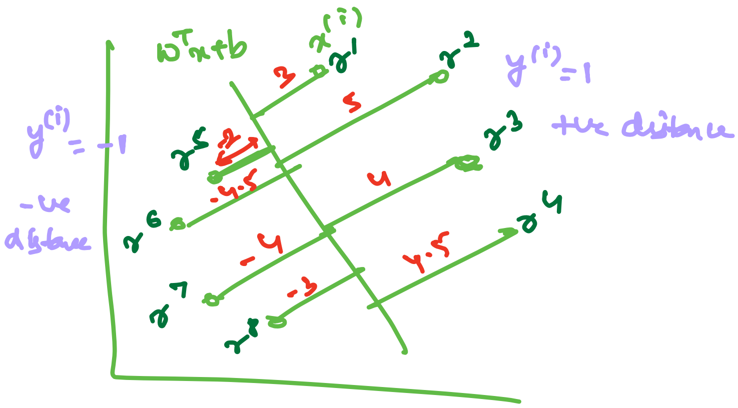
$$g(z) = +1 \quad \text{if } z \geq 0$$

class +ve

$$g(z) = -1 \quad \text{if } z < 0$$

class -ve

Loss fx^n:



distance of ith point from d.b.

$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{\|w\|_2}$$

$$\gamma = \boxed{\min_{i=1 \dots m} \gamma^{(i)}}$$ → among all the distances find the minimum

target: $\max \gamma$

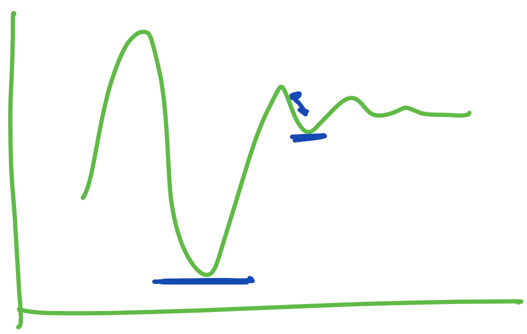
SVM objective

$$\begin{cases} \max \gamma \\ \text{such that } \dfrac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|_2} \geq \gamma \quad \text{for all } i = 1 \dots m \end{cases}$$

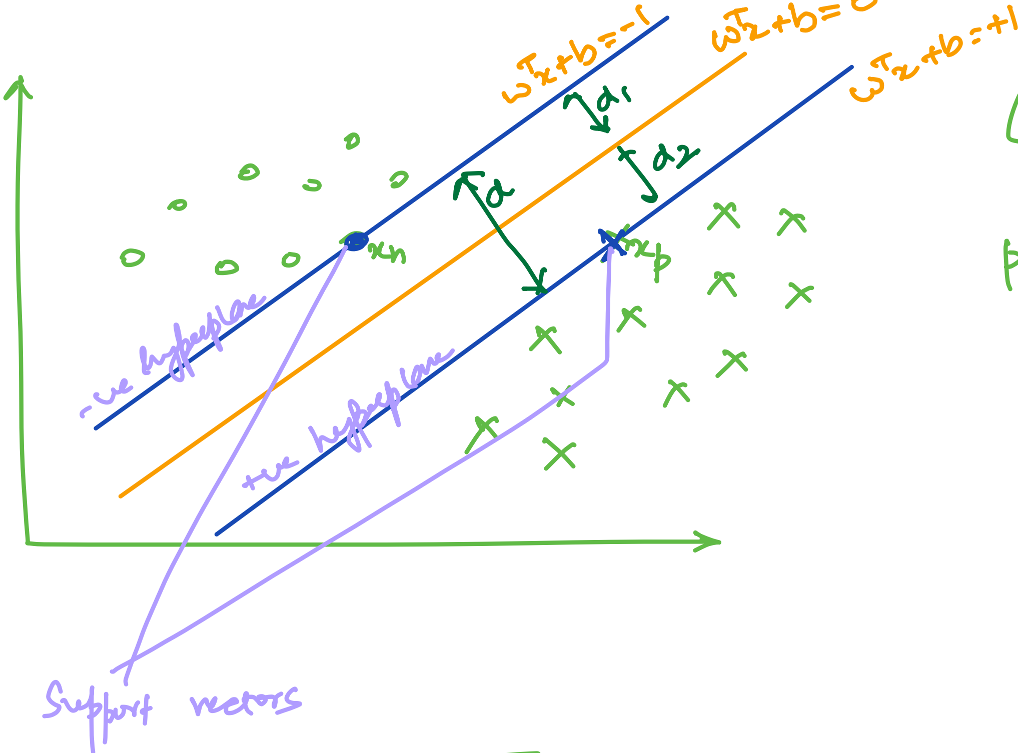
$\dfrac{w^T x^{(i)} + b}{\|w\|_2}$ → normalized distance +ve / -ve

$\dfrac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|_2}$ → normalized absolute distance

SVM objective → non convex



Re-formulate the SVM objective:



(Re-normalize our data points such that points which are closest to hyperplane have distance +1 or -1)

$$d_1 = \frac{|w^T x_n + b|}{\|w\|_2} \qquad d_2 = \frac{\|w^T x_p + b\|}{\|w\|_2}$$

$w^Tx+b=-1$ and x_n lies in it

$$w^T x_n + b = -1 \qquad w^T x_p + b = 1$$

$$d_1 = \frac{1}{\|w\|_2} \qquad d_2 = \frac{1}{\|w\|_2}$$

$$d = d_1 + d_2 = \frac{2}{\|w\|_2}$$

maximize distance d

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

SVM objective:

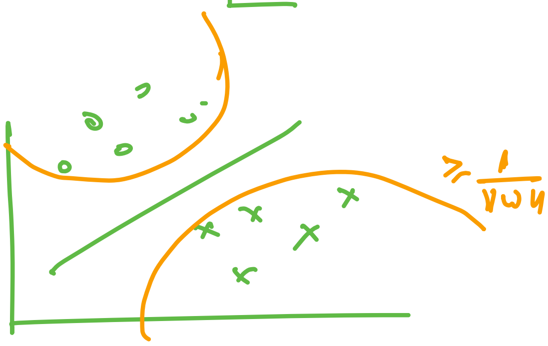
$$\min \frac{\|w\|}{2}$$

under the condition all points should have min distance ≥

$$\frac{y^{(i)}(w^T x^{(i)}+b)}{\|w\|_2} \geq \frac{1}{\|w\|_2}$$

SVM objective

$$\left[\begin{array}{l} \min \frac{\|w\|_2}{2} \\ \text{such that } \quad y^{(i)}(w^T x^{(i)}+b) \geq 1 \\ \forall\, i \in \{1 \ldots m\} \end{array}\right.$$



all data points should have atleast distance of $\frac{1}{\|w\|}$

$$w = \sqrt{w_1^2+w_2^2} \quad \text{or} \quad \left(\sum w_{(i)}^2\right)^{1/2}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \qquad w^T = [w_1, w_2 \ldots w_n]$$

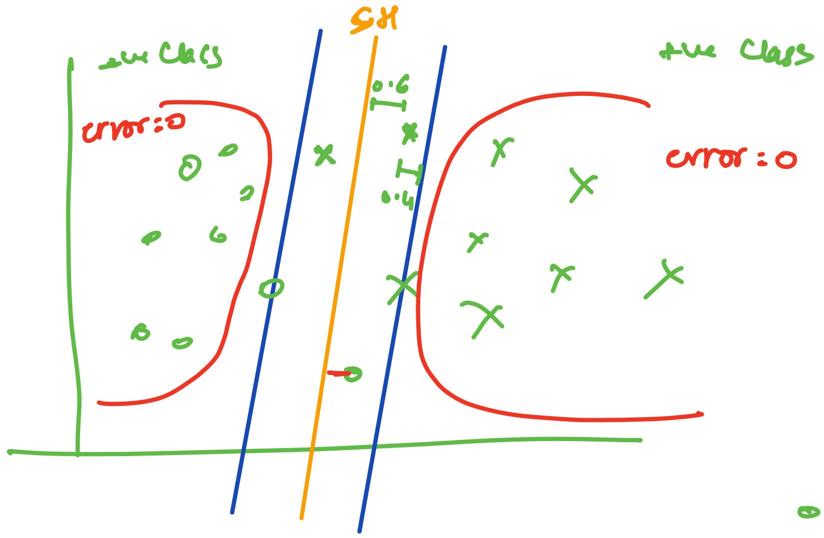
$$\underline{ww^T} \Leftarrow w^2 = w_1^2+w_2^2 \longrightarrow 2w_1+2w_2 = 2(w_1+w_2)$$

Final SVM objective

$$\left.\left[\begin{array}{l} \min \frac{\|w\|_2^2}{2} \\ \text{such that } \quad y^{(i)}(w^T x^{(i)}+b) \geq 1 \\ \forall\, i \in \{1 \ldots m\} \end{array}\right.\right\}$$

convex fxn with some constraints.

Handle Errors



$$y^{(i)}(w^T x^{(i)}+b) \geq 1$$

$$\underbrace{y^{(i)}(w^T x^{(i)}+b)}_{\text{distance of } x^{(i)} \text{ from SH.}} \geq \underbrace{\overbrace{1-\varepsilon^{(i)}}^{0.6}}_{\text{Allow some errors}} \longrightarrow 0.4$$

$\varepsilon^{(i)}$ denotes the distance of ith point from the hyperplane

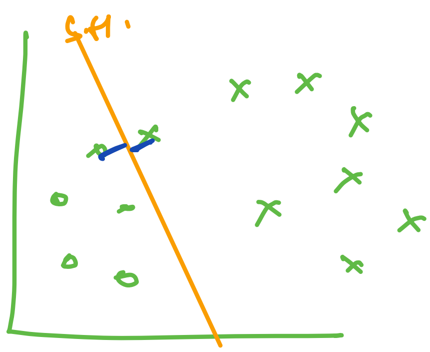
$$\text{loss} = \min \frac{1}{2} w \cdot w^T + \overset{\text{constant}}{C} \sum_{i=1}^{m} \varepsilon^{(i)}$$

$$\text{s.t.} \quad y^{(i)}(w^T x^{(i)}+b) \geq 1-\varepsilon^{(i)}$$

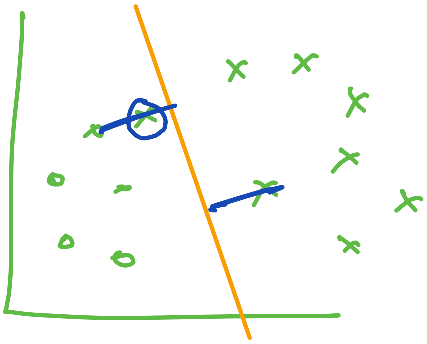
C = hyperparameter

$C=\infty$ 0 Errors

you won't be able to achieve maximum margin's hyperplane



C= 1 afford some errors, maximum margin sep. hyperplane



Non Convex

↓

Convex

↓

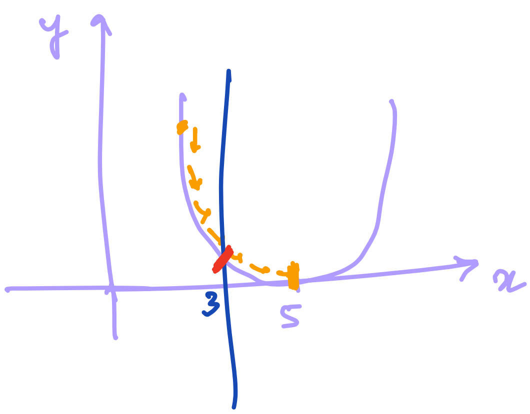
Misclassification Errors.

$\boxed{\text{distance } 1}$

$y = (x-5)^2$ value of x for which y is minimum. → no constraint

5

$y = (x-5)^2$ such that $x \le 3$ → constraints



$$\text{Loss} = \min \frac{1}{2} w \cdot w^T + \overset{\text{constant}}{C} \sum_{i=1}^{m} \varepsilon^{(i)}$$

$$\text{st.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \varepsilon^{(i)}$$

Convex fx^n with constraints: we want to get rid of linear constraints.

Remove linear constraints

$$\varepsilon^{(i)} \geq 1 - \boxed{y^{(i)}(w^T x^{(i)} + b)} \quad t^{(i)}$$

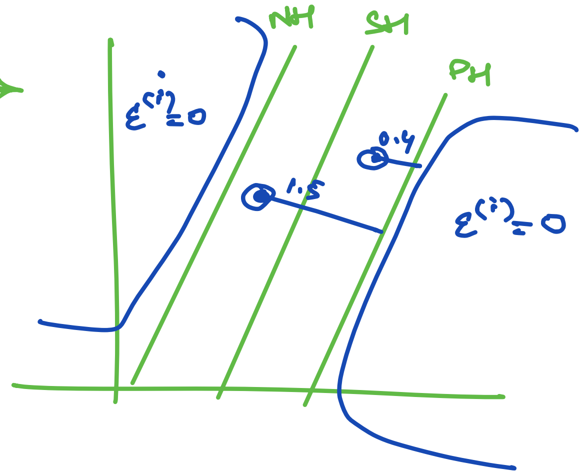
unnormalized absolute distance of $x^{(i)}$ from separating hyperplane.

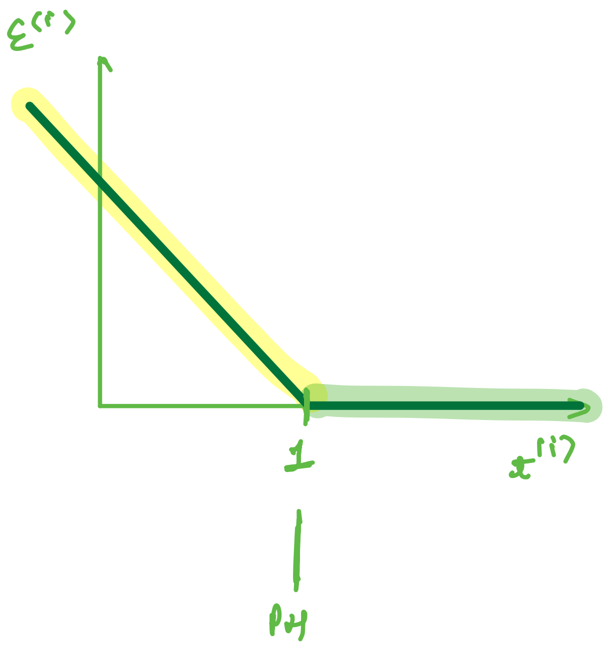
$$\varepsilon^{(i)} \geq 1 - t^{(i)}$$

$$\varepsilon^{(i)} \geq 0$$

combine

$$\varepsilon^{(i)} = \max(0, 1 - t^{(i)})$$





for differentiating $\varepsilon^{(i)}$, we will use the concept of subgradient

$$\begin{cases} 0 & \text{if } t_i \geq 1 \\ -1 & \text{if } t_i < 1 \end{cases}$$

$$\boxed{L = \frac{1}{2} w^T w + C \sum_{i=1}^{m} \max(0, 1 - t^{(i)}) \quad \text{where } t^{(i)} = y^{(i)}(w^T x^{(i)} + b)}$$

SVM Objective

SH

$\boxed{w_1\ w_2 \ldots w_n\ b\ ?}$

-ve +ve

GRADIENT DESCENT

$$w = w - \eta \frac{\partial L}{\partial w} \rightarrow \nabla_w L$$

$$\frac{1}{2} w^T \cdot w = \frac{1}{2}(w_1^2 + w_2^2 + \dots w_n^2)$$

$$\frac{\partial}{\partial w}\left(\frac{1}{2} w^T w\right) = \frac{1}{2}(2w_1 + 2w_2 + \dots 2w_n) = \boxed{w_1 + w_2 + \dots w_n} \to w$$

$$\frac{\partial L}{\partial w} = w + c \sum_{i=1}^{m} \frac{\partial}{\partial w} \underbrace{\left(\max(0, 1 - t^{(i)})\right)}_{f}$$

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial t_i} \frac{\partial t_i}{\partial w}$$

$$\frac{\partial L}{\partial w} = w + c \sum_{i=1}^{m} \frac{\partial f}{\partial t^{(i)}} \cdot \frac{\partial t^{(i)}}{\partial w}$$

$$\frac{\partial L}{\partial w} = w + c \sum_{i=1}^{m} \frac{\partial}{\partial t_i}\left(\max(0, 1 - t^{(i)})\right) \cdot \frac{\partial t^{(i)}}{\partial w}$$

$$\frac{\partial L}{\partial w} = w + c \sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if} & t^{(i)} \geq 1 \\ -1 & \text{if} & t^{(i)} < 1 \end{bmatrix} \frac{\partial t^{(i)}}{\partial w}$$

$$t^{(i)} = y^{(i)}(w^T x^{(i)} + b)$$

$$\frac{\partial t^{(i)}}{\partial w} = y^{(i)} x^{(i)}$$

$$\boxed{\frac{\partial L}{\partial w} = w + c \sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if} & t^{(i)} \geq 1 \\ -1 & \text{if} & t^{(i)} < 1 \end{bmatrix} y^{(i)} x^{(i)}}$$

$$b = b - \eta \frac{\partial L}{\partial b} \;?$$

$$\frac{\partial L}{\partial b} = 0 + c\sum_{i=1}^{m} \frac{\partial}{\partial b}\Big[\underbrace{\max(0, 1 - t_i)}_{f}\Big]$$

$$\frac{\partial L}{\partial b} = c\sum_{i=1}^{m} \frac{\partial f}{\partial t_i} \cdot \frac{\partial t^{(i)}}{\partial b}$$

$$\frac{\partial L}{\partial b} = c\sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ -1 & \text{if } t^{(i)} < 1 \end{bmatrix} \frac{\partial t^{(i)}}{\partial b}$$

$$\frac{\partial t^{(i)}}{\partial b} = \frac{\partial}{\partial b}\left(y^{(i)}(w^T x^{(i)} + b)\right)$$

$$= y^{(i)}$$

$$\boxed{\frac{\partial L}{\partial b} = c\sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ -1 & \text{if } t^{(i)} < 1 \end{bmatrix} y^{(i)}}$$

UPDATE RULES:

$$w = w - \eta\left[w + c\sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ -1 & \text{if } t^{(i)} < 1 \end{bmatrix} y^{(i)} x^{(i)}\right]$$

$$\boxed{w = w - \eta w + \sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ \eta c y^{(i)} x^{(i)} & \text{if } t^{(i)} < 1 \end{bmatrix}}$$

$$b = b - \left(\eta c \sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ -1 & \text{if } t^{(i)} < 1 \end{bmatrix} y^{(i)} \right)$$

$$\boxed{b = b + \sum_{i=1}^{m} \begin{bmatrix} 0 & \text{if } t^{(i)} \geq 1 \\ \eta c y^{(i)} & \text{if } t^{(i)} < 1 \end{bmatrix}}$$