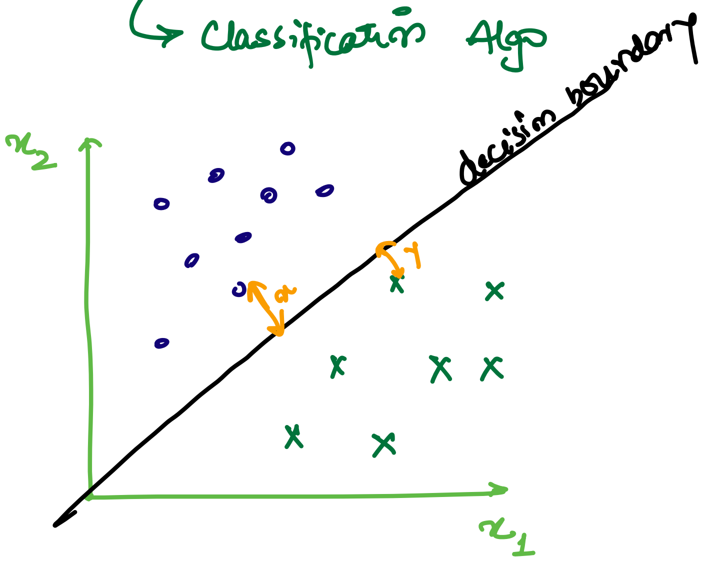
SVM (Support Vector Machine)

↳ Classification Algo

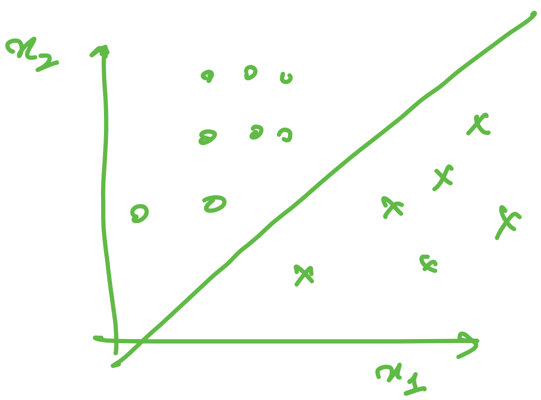


Points which are closest to your decision boundary should be very far away from each other.

Hyperplane:

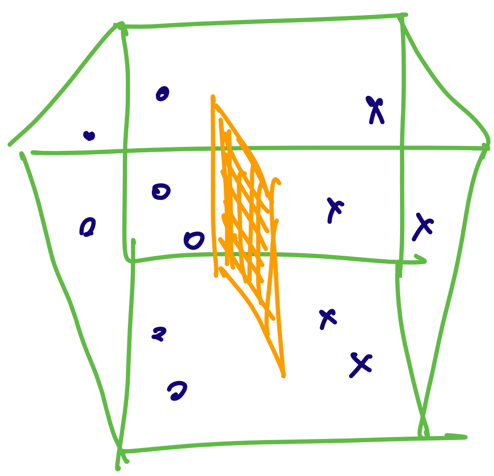
n features

hyperplane n-1 dimension



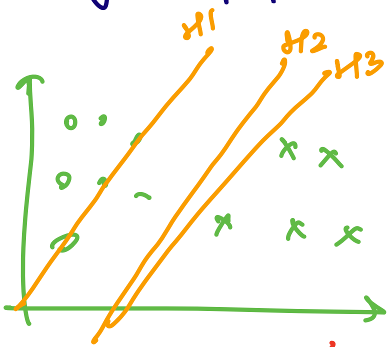
→ line (1D)

3 features



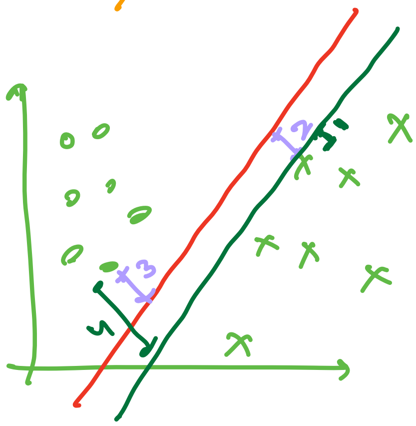
2D hyperplane: plane

Separating Hyperplane:



$H_2, H_3 \rightarrow$ Separating Hyperplane

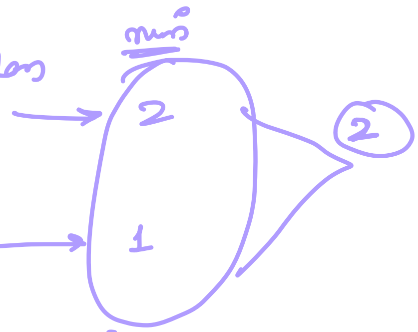
H_1 ✗



min distance from each class

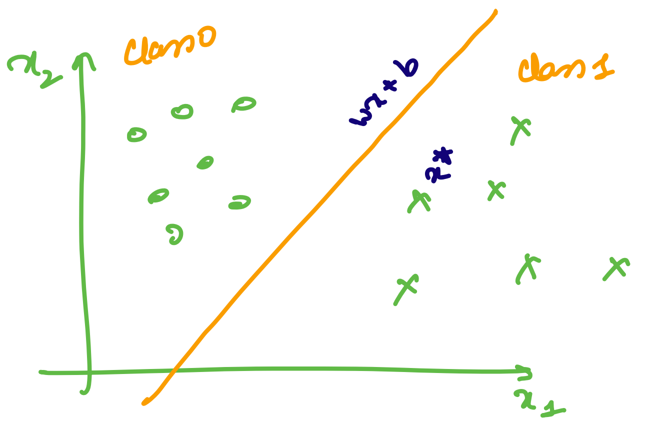
(2, 3)

(1, 4)



min distance of dp from sep. hyperplane

maximize the minimum distance from s.h.



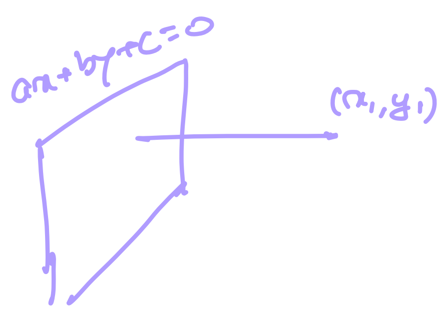
$y = mx + c$

$y = wx + b$

$wx^* + b$

$wx^* + b > 0$ Class 1

$wx^* + b < 0$ Class 0



$$\frac{ax_1 + by_1 + c}{\sqrt{a^2 + b^2}}$$

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

L_2 norm: $(a^2 + b^2)^{1/2}$

⋮

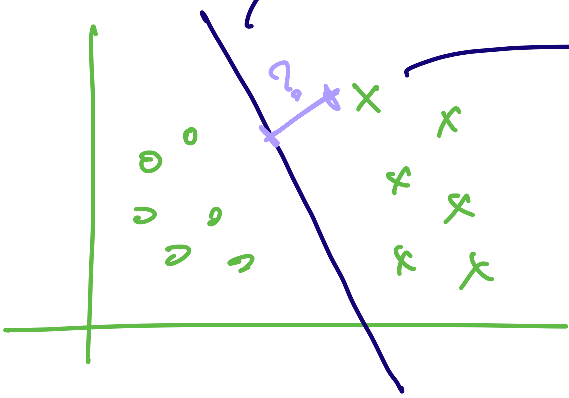
L_3 norm $(a^2+b^3)^{1/3}$

$ax+b \rightarrow wx+b \longrightarrow w_1x_1 + w_2x_2 + b = 0$

b: bias

x_1, x_2 features

w_1, w_2 weights

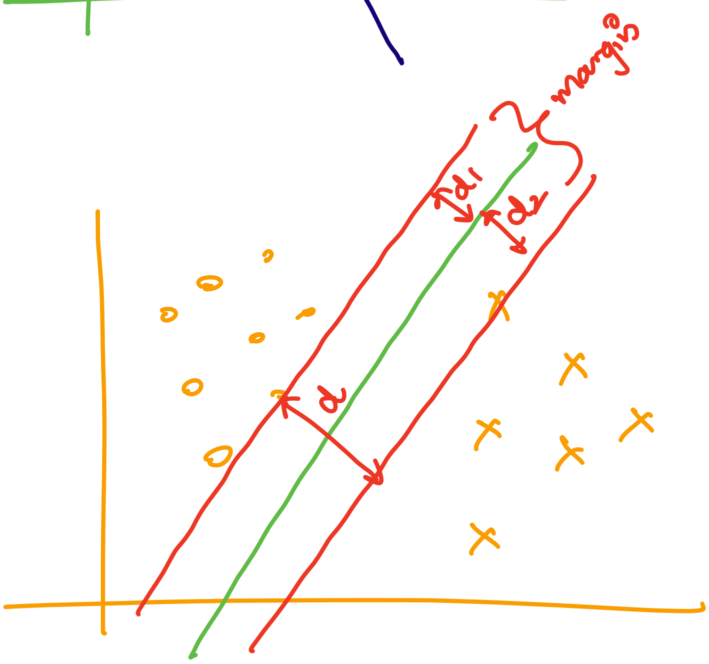


$x_1^{(i)}, x_2^{(i)}$

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

L_2 norm

$\|w\|_2$



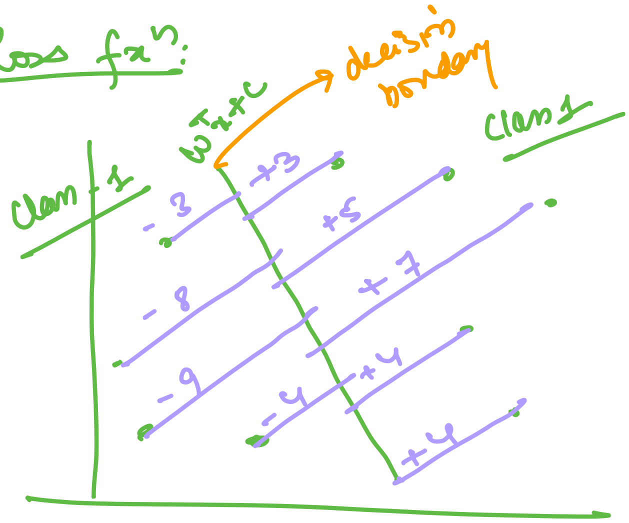
formulating objective:

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

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

Binary Classification $y^{(i)} \in \{-1, 1\}$ → class labels

Loss fx^n:



$$\gamma^{(i)} = \frac{w^T x^{(i)} + b}{\|w\|_2}$$ → distance of the point from decision boundary

$$\gamma = \min_{i=1\ldots m} \gamma^{(i)}$$ ③

target: max γ

$$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \vdots \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \end{bmatrix}$$

$$w^T x = \begin{bmatrix} w_1 & w_2 & w_3 & w_4 & \ldots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

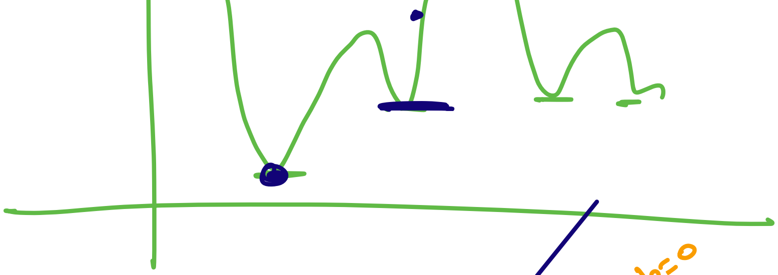
$$= w_1 x_1 + w_2 x_2 + w_3 x_3 \ldots$$

SVM objective
$$\left[\begin{array}{l} \max \gamma \\ \text{such that } \dfrac{y^{(i)}(w^T x^{(i)} + b)}{\|w\|} \geq \gamma \quad \text{for all } i = 1 \ldots m \end{array}\right.$$

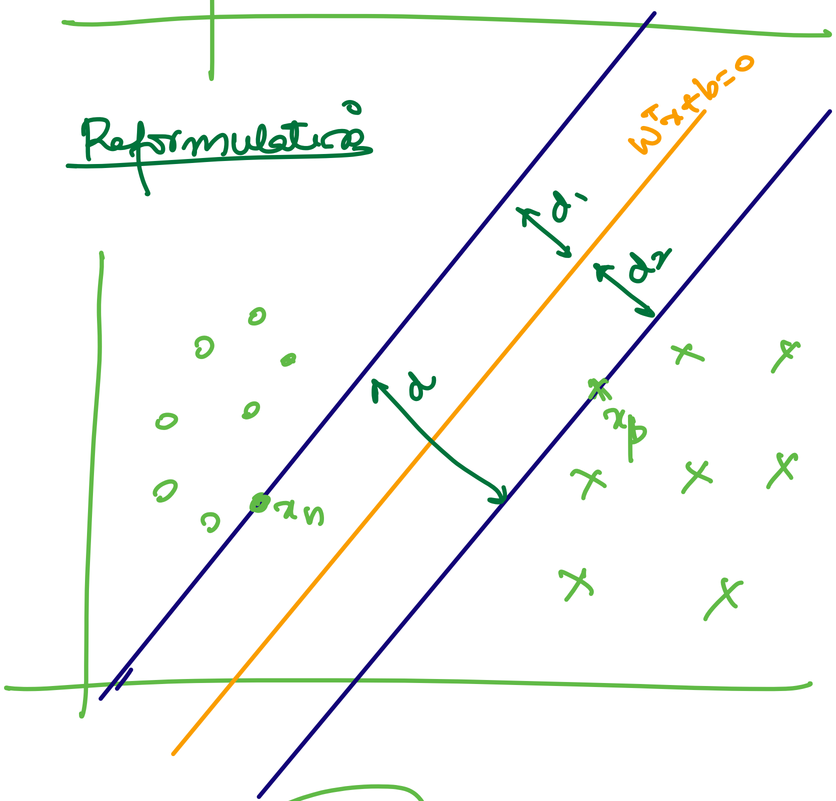
$$\frac{w^T x^{(i)} + b}{\|w\|_2}$$ → normalized distance → +, → −

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

SVM objective → non convex



Reformulation



Re-normalize our data points such that point which is closest to hyperplane lies at distance +1, -1

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

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

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

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

$$\max d \Rightarrow \min \frac{\|w\|_2}{2}$$

SVM objective: $\min \frac{\|w\|}{2}$

under the condition that all points should have min distance 1.

$$\frac{y^{(i)}(w^T x^{(i)}+b)}{\|w\|_2} \geq \frac{1}{\|w\|_2} \qquad d_1 = \frac{1}{\|w\|_2}$$

SVM Objective:

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

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

$$\|w\|_2 = \sqrt{w_1^2+w_2^2}$$

$$\|w\|_2^2 = w_1^2+w_2^2 = w \cdot w^T$$

differentiation

$$2w_1+2w_2=0$$

$$w_1+w_2=0$$

$$w^T w = \begin{array}{cc} w^T & w \\ [w_1 \; w_2 \; \cdots \; w_n] & \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \end{array}$$

$$= w_1^2 + w_2^2 + \cdots w_n^2$$

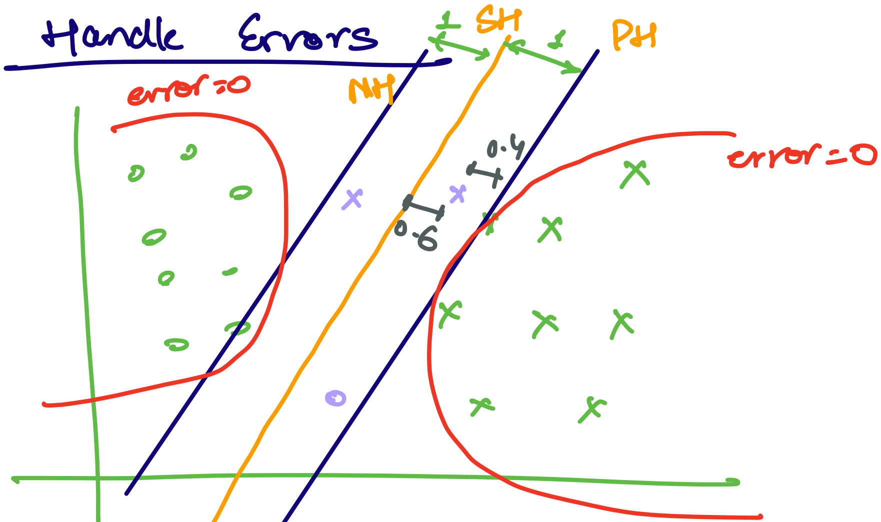
SVM Objective:

$$\min \frac{\|w\|_2^2}{2}$$

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

$\forall \; i \in \{1, \ldots m\}$

Handle Errors



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

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

$y^{(i)}(w^T x^{(i)} + b) \geq \underbrace{1 - \varepsilon^{(i)}}_{0.6} \rightarrow 0.4$

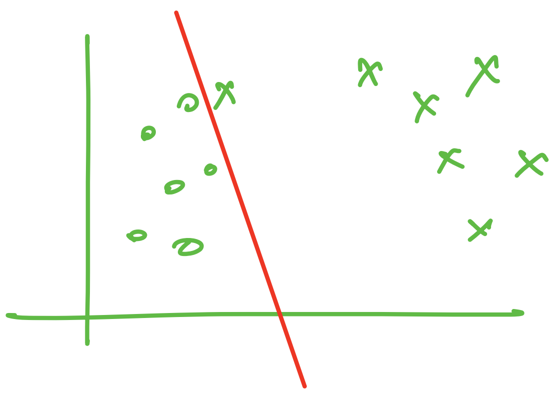
Allowing some errors

Loss : $\left[\begin{array}{l} \min \; \frac{1}{2} \overbrace{w^T \cdot w}^{\frac{\|w\|^2}{2}} + C \sum_{i=1}^{m} \varepsilon^{(i)} \\ \text{such that } \; y^{(i)}(w^T x^{(i)} + b) \geq 1 - \varepsilon^{(i)} \end{array} \right.$

C = hyperparameter

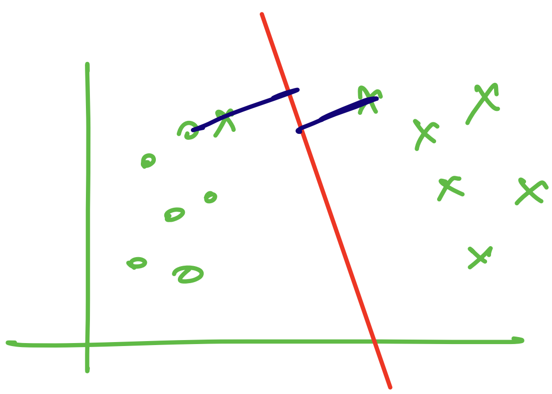
$C = \infty$ No Errors

$\min \left[\_\_\_ + 1000 \cdot \_\_\_ \right]$



You will not be able to achieve maximum margin hyperplane.

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



Non Convex

↓

Convex $\quad$ distance = 1

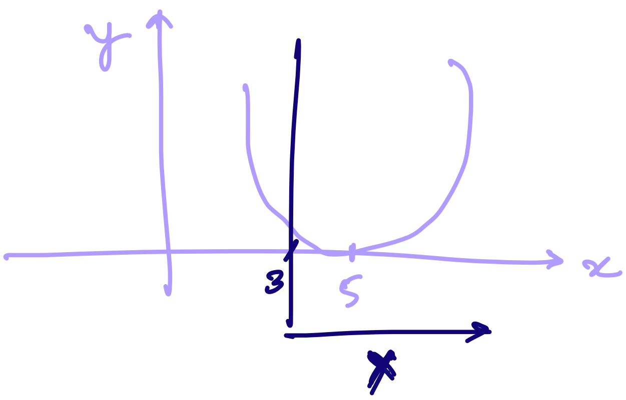
↓

Errors

$y = (x-5)^2$

Value of x for which y is min?

↳ $\leq$



such that $x \leq 3$ $\quad$ Constraint

Remove linear constraints

Loss :

$$\left[\begin{array}{l} \min \; \frac{1}{2} w^T w + C \sum_{i=1}^{m} \varepsilon^{(i)} \\ \text{such that } \; y^{(i)}(w^T x^{(i)} + b) \geq 1 - \varepsilon^{(i)} \end{array}\right]$$

Convex fn constraints : we want to remove constraints.

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

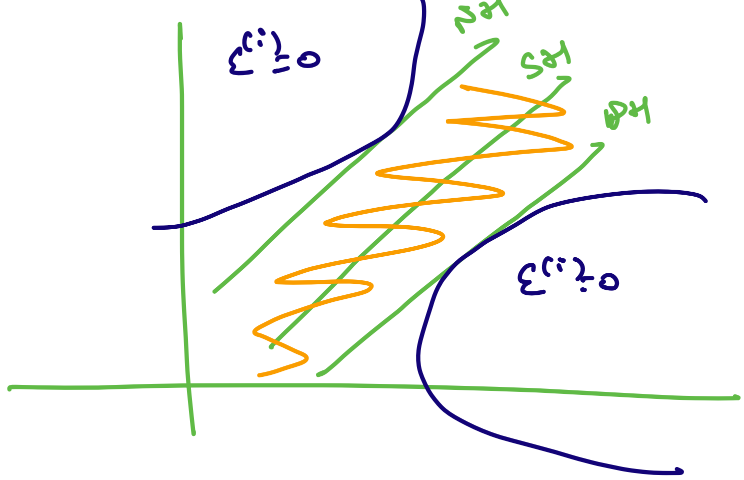
$t^{(i)}$ unnormalized absolute distance of $x^{(i)}$ from s.h.

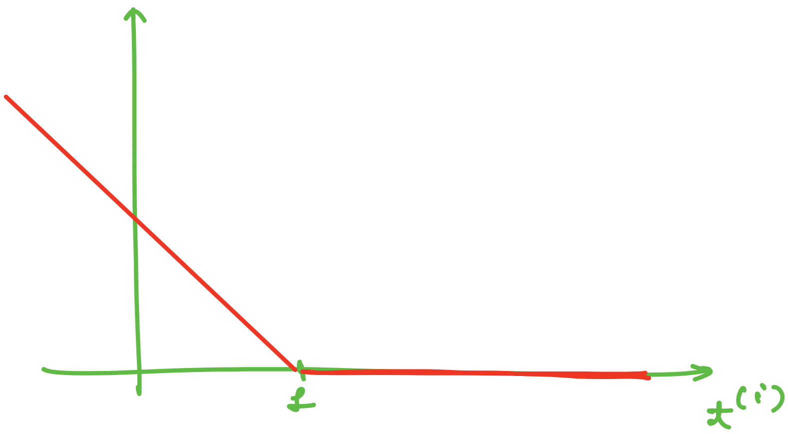
$$\varepsilon^{(i)} \geq 1 - t^{(i)}$$

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

$\big\downarrow$ combine

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





for differentiating $\epsilon^{(i)}$

concept of subgradient

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

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

SVM objective

Random value of w

How good present w is?

update w's → Gradient Descent

$w = w - \eta \frac{\partial L}{\partial w}$?

$w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ \vdots \end{bmatrix}$ $\quad w^T = [w_1 \; w_2 \ldots]$

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

$= w_1^2 + w_2^2 \ldots - - w_n^2$

$= ||w||_2^2$

$$b = b - \eta \left(\frac{\partial L}{\partial b}\right) \,?$$

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

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

$$= \left(w_1 + w_2 + \dots \; w_n\right) = w$$

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

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial t} \cdot \frac{\partial t}{\partial w}$$

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

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

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

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

$$\frac{\partial t^{(i)}}{\partial w_1} = y^{(i)} x^{(i)}$$

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

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

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

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

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

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

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

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

<u>UPDATE RULES:</u>

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

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

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

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