Loss Minimization in Classification

Agenda

⇒ Recap

  → Distance b/w origin and Line

  → Distance b/w point and Line

⇒ Distance b/w 2 planes

⇒ Putting it all together: loss function

  → Gain function: Maths

  → Convert gain function to loss function

⇒ Weigth update:

  → The perceptron Algo

  → Coding the Perceptron

⇒ Non-Linear Decision Boundary: Circle

Recap

a) Vectors : $\bar{x}$ $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix}$ $\bar{x} \in \mathbb{R}^d$

$X_{matrix} \in \mathbb{R}^{n \times m}$

b) Norm : $||x||$

$l2 \text{ Norm} = \sqrt{x_1^2 + x_2^2 \text{---} x_d^2}$

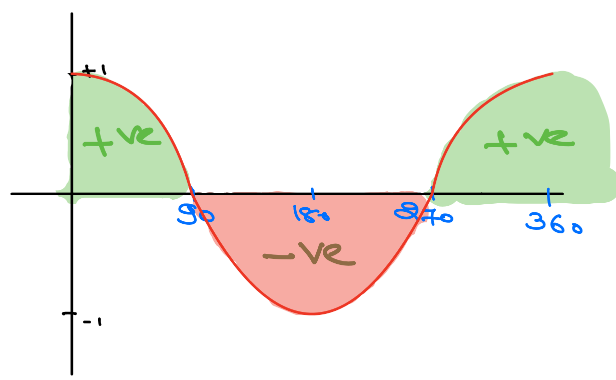
$l1 \text{ Norm} = |x_1| + |x_2| \text{-----} |x_d|$

c) Dot product :

$\bar{x} \cdot \bar{y} \Rightarrow x^T y \Rightarrow x_1 y_1 + x_2 y_2 \text{---} x_d y_d$

d) Angle b/w Vectors :

$$\cos\theta = \frac{\bar{x}^T y}{||x|| \, ||y||}$$

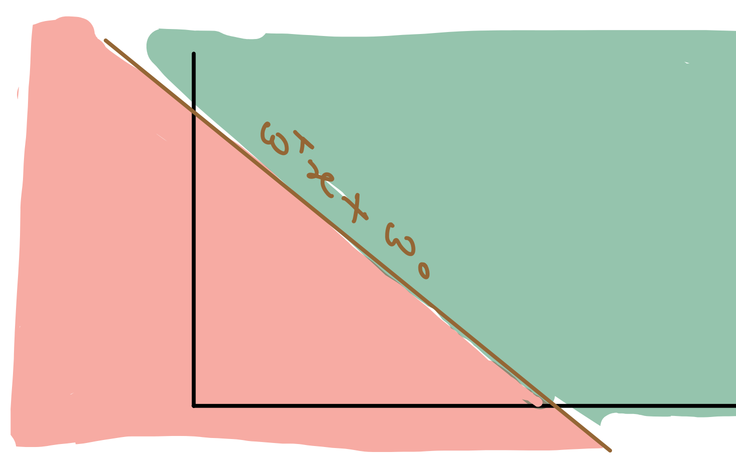


f) Eqn of line and HalfSpaceces

$w_1 x_1 + w_2 x_2 \text{---} w_0 = 0$

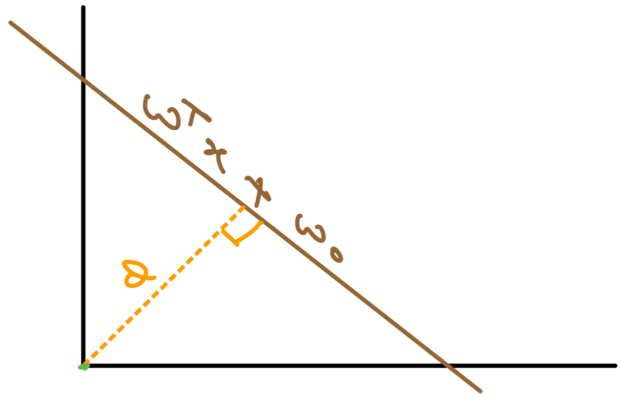
$\downarrow$

$w^T X + w_0 = 0$



e) Dist b/w origin and Line:

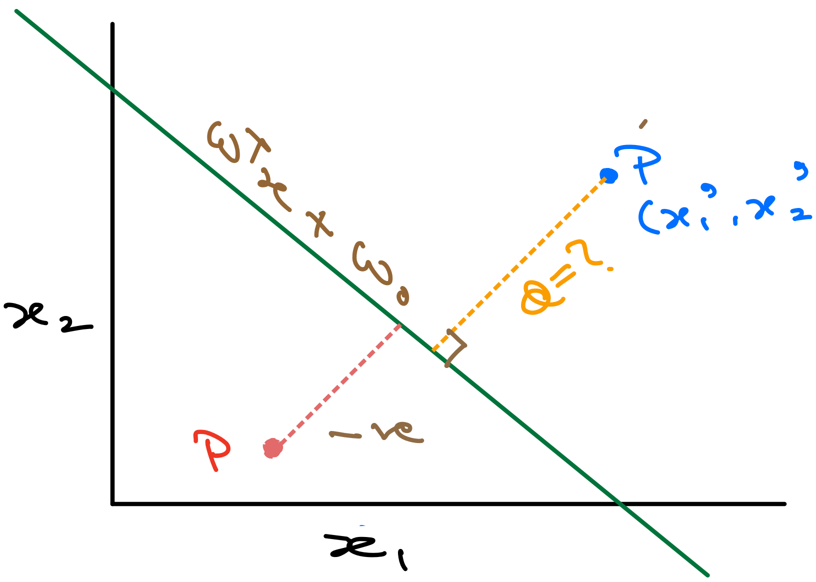
$$\frac{w_0}{\|\vec{w}\|}$$



f) Dist b/w point and Line

$$\boxed{d \Rightarrow \frac{w^T p + w_0}{\|w\|}}$$

+ve



$$w^T x + w^0$$

$$\downarrow$$

$$w^T p + w_0 \begin{cases} \nearrow +ve \\ \searrow -ve \end{cases}$$

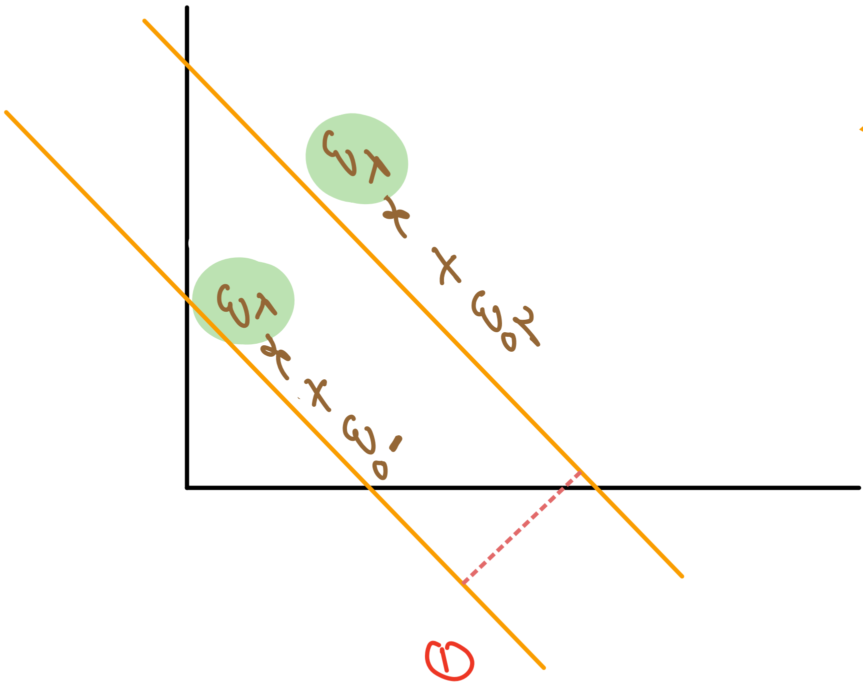
Q) Eq^n $\Rightarrow$ $2x - 3y + 6z = 2$ point = $(-6, 0, 0)$

which HF point lies in?

$$w^T x_i + w_0 \longrightarrow 2 \times (-6) - 3 \times 0 + 6 \times 0 - 2$$

$$-12 - 2$$

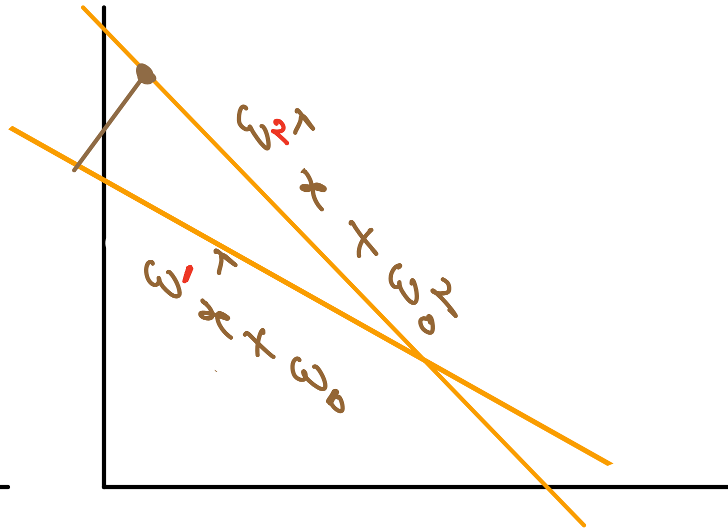
Distance b/w 2 Hyperplanes



(1)

$w_1 = w_2 = w^T$

dist != 0



(2)

$w_1 \neq w_2$

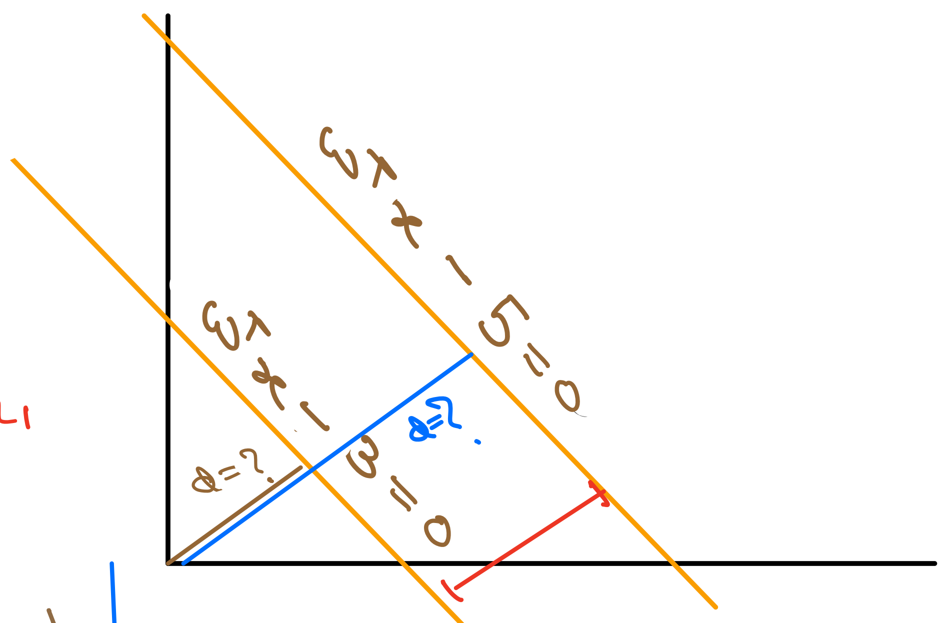
dist = 0

$d_{0 \to L2} \Rightarrow \frac{5}{||w||}$

$d_0 \to L_2 = \frac{3}{||w||}$

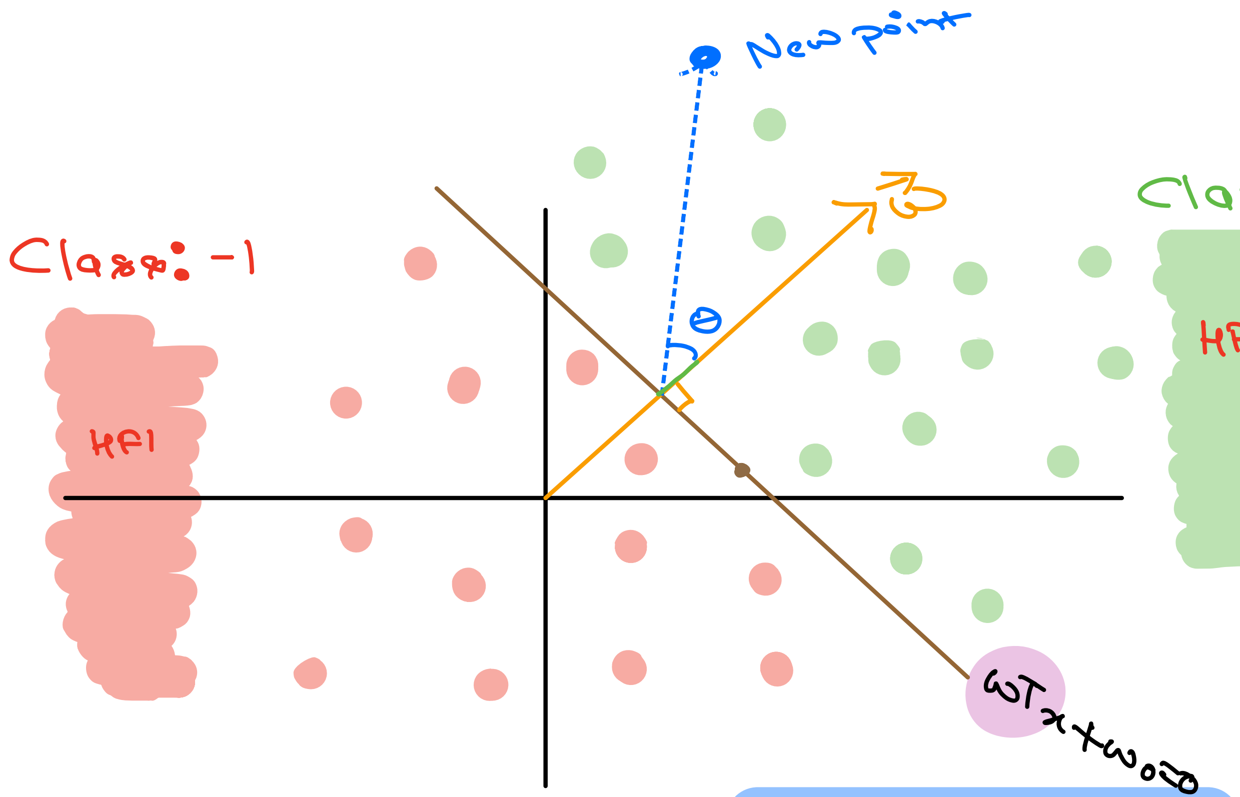
$d_{L2 \to L1} = d_{0 \to L2} - d_{0 \to L1}$

$d_{planes} = \left| \frac{w_0^2 - w_0^1}{||w||} \right|$



Ans = 2 / ||w||

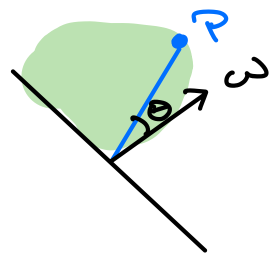
HalfSpace a Point belongs to?



$p \Rightarrow [x_1', x_2']$?

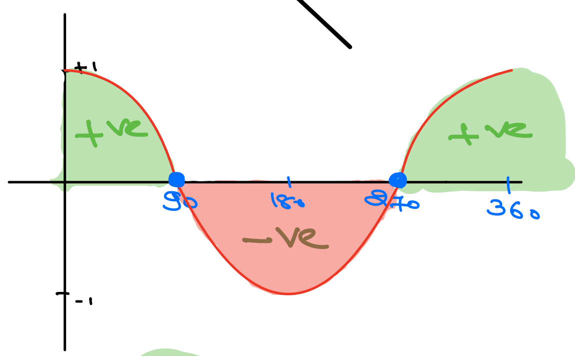
① $0 < \theta < 90$

$\cos\theta = +ve = w^T p$



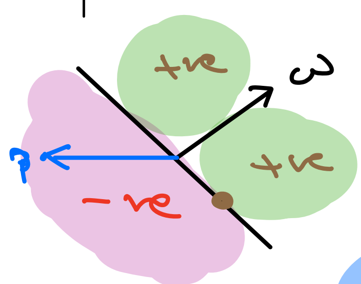
② $90 < \theta < 180$

$\cos\theta = -ve = w^T p$



③ $180 < \theta < 270$

$\cos\theta = -ve = w^T p$



④ $270 < \theta < 360$

$\cos\theta = +ve = w^T p$

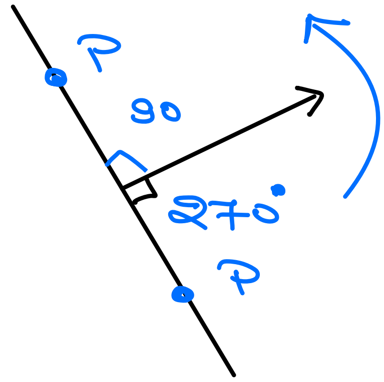
$$\cos\theta \Rightarrow \frac{a \cdot b}{\|a\|\,\|b\|} \longrightarrow \cos\theta \Rightarrow \frac{w^T p}{\|w\|\,\|p\|}$$

+ve

+ve

-ve

① $\cos\theta$ is +ve when $\vec{w}^T p > 0$
② $\cos\theta$ is −ve when $\vec{w}^T p < 0$
③ $\cos\theta$ is 0 when $\vec{w}^T p = 0$

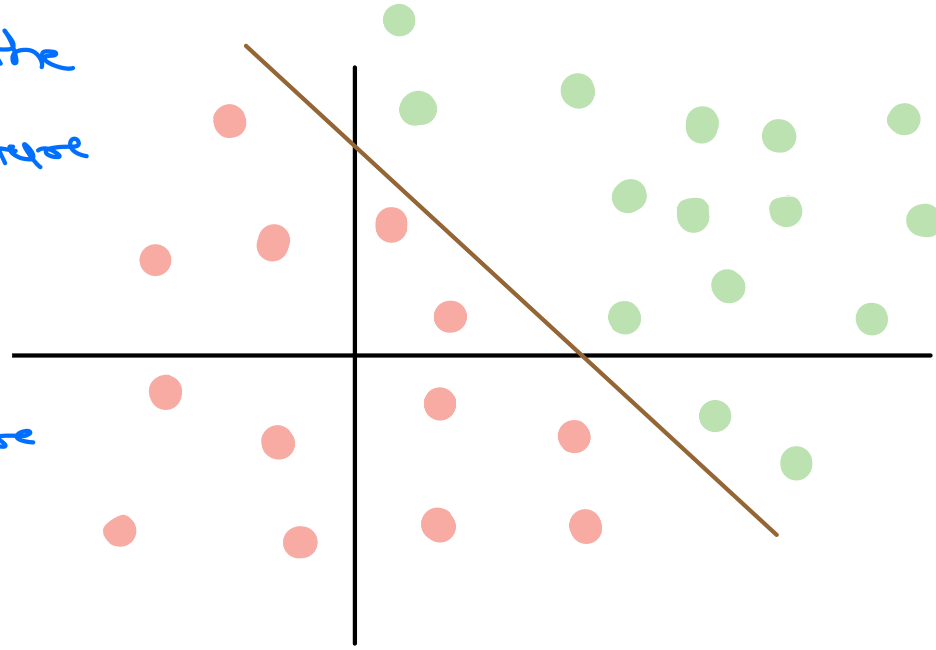


$$\cos\theta = 0 \Rightarrow w^T p = 0$$

Putting it all Together
Loss Function

① Loss function tells us how good the geometric structure

② Lower the Loss better the structure

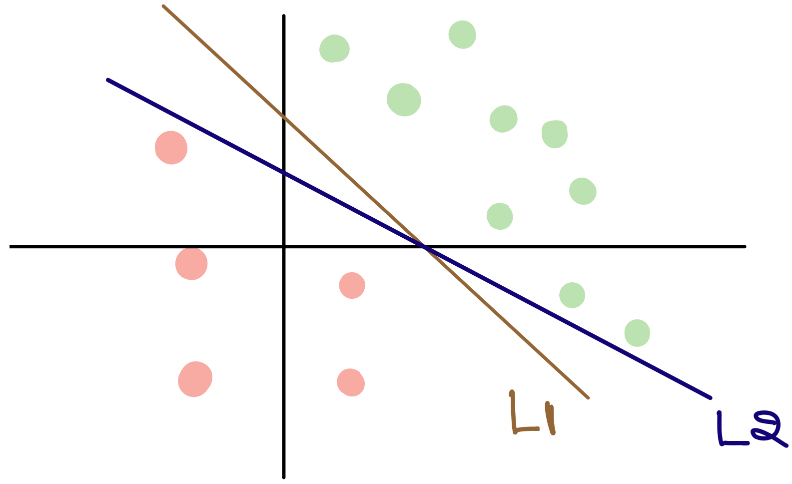


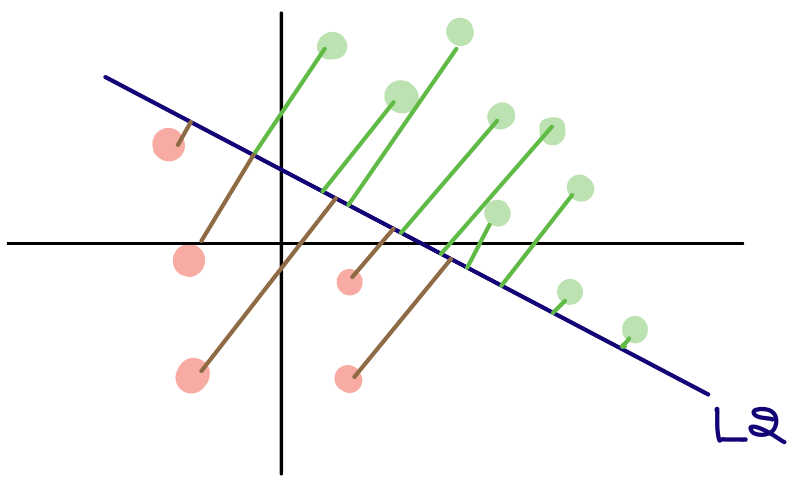
① How are labels assigned to New Datapoint

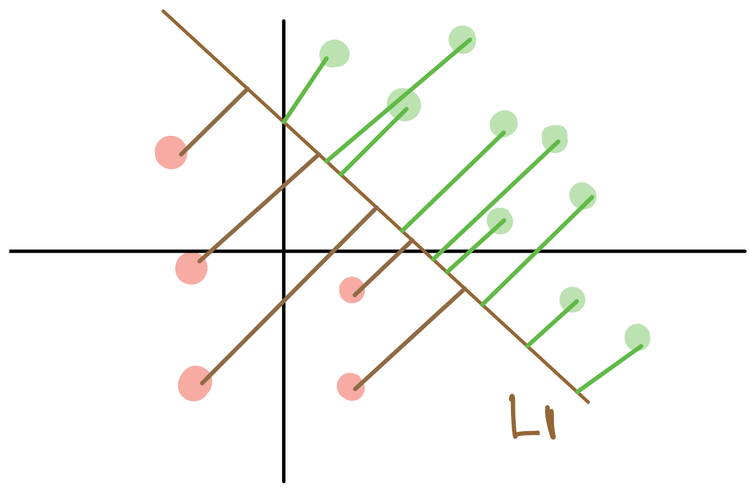
$$d = \frac{w^T x + w_0}{\|w\|} \rightarrow \text{sign}(w^T p + w_0)$$

+ve = +1   −ve = −1

Q Which line is better : L1 or L2?







Avg(L1_D) > Avg(L2_D)

sum(L1_D) > sum(L2_D) ✓ (Perceptron)

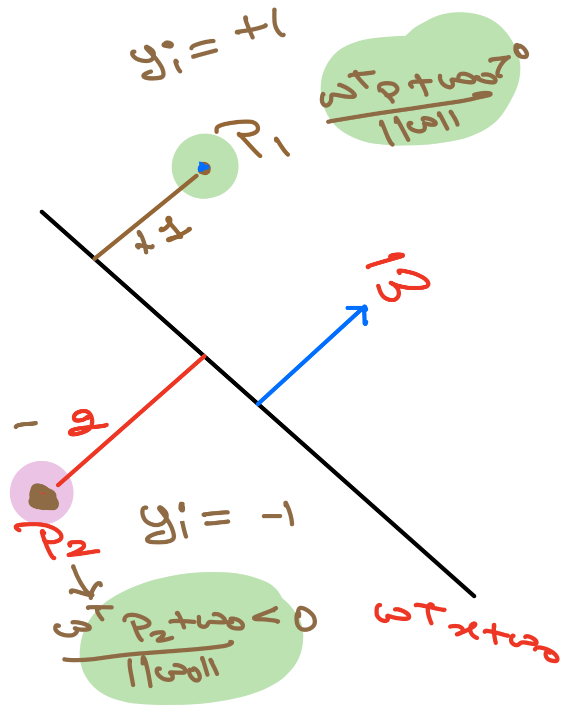
min(L1_D) > min(L2_D) (SVM)

* A Good Line will have S.O.D as high as possible

Gain func $\Rightarrow$ maxi $\left(\sum_{i=1}^{n} \frac{w^T x_i + w_0}{\|w\|} \right)$

Loss $f(x) \Rightarrow -$ Gain $f(x)$

$$\left.\begin{matrix} w_1 \\ w_2 \\ \vdots \\ w_{1000} \end{matrix}\right\} \quad \begin{matrix} \text{Max } G(x) \\ \text{least } L(x) \end{matrix}$$



1. Issue with current G and L function
2. How do we rotate vectors

$P_1 \Rightarrow +ve$ $\quad P_2 \Rightarrow -ve$

$$\text{Gain func} \Rightarrow \text{maxi} \left(\sum_{i=1}^{n} \frac{w^T x_i + w_0}{\|w\|} \right) \times y_i$$

How do we fix this issue

1. Take squares ⇒ ✓ $\quad 1^2 + (-2)^2 \Rightarrow 5$
   (Calculation is complex) $\quad \sqrt{5}$
2. Take absolute ⇒ ✓ $\rightarrow |1| + |-2| = 3$
   (Derivative @ 0 X)
3. Multiply with y_i ⇒ $(1) \times 1 + (-2) \times (-1)$ ✓
   (Easy to calculate) $\Rightarrow 3$

Mathematical Notation

① Dataset : $X = \{(\vec{x}_i, y_i)\}_{i=1}^{n}$

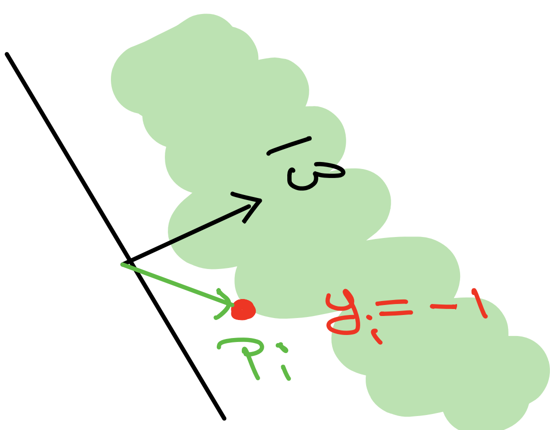
② Classifier or Model : $\vec{w}^T \vec{x} + w_0 = 0$

③ Loss function : – Gain Function

– (Sum of distances b/w CLF and $\vec{x}$)

$$-\sum_{i=1}^{n} \left(\frac{\vec{w}^T \vec{x}_i + w_0}{\|\vec{w}\|} \right) \times y_i$$

Goal : minimize $_{\vec{w},\, w_0}$ $-\sum_{i=1}^{n} \left(\frac{\vec{w}^T \vec{x}_i + w_0}{\|\vec{w}\|} \right) \times y_i$



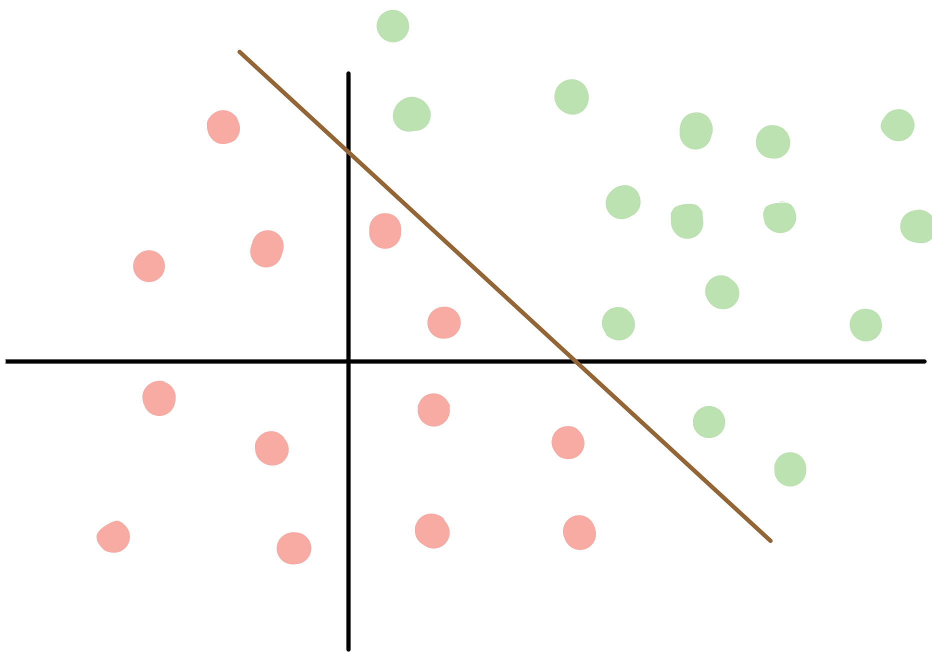
$\hat{y}_{P_i} \Rightarrow +1$

misclassification $\Rightarrow$ $\hat{y}_i \;!=\; y_i$

Pred != Actual_label

How do we reduce Misclassification?

Perceptron Learning Algorithm



Q1: How do we identify that a point is misclassified?

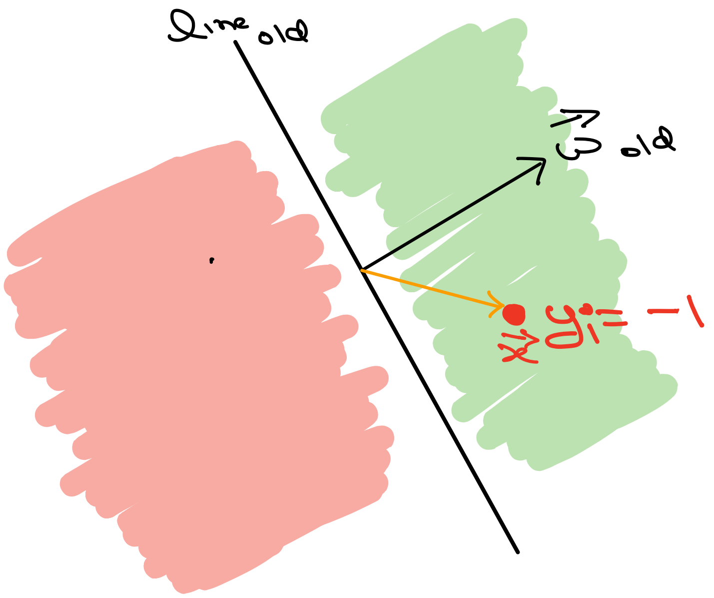
if $sign(y_i) == sign(W^T x + w_0)$:

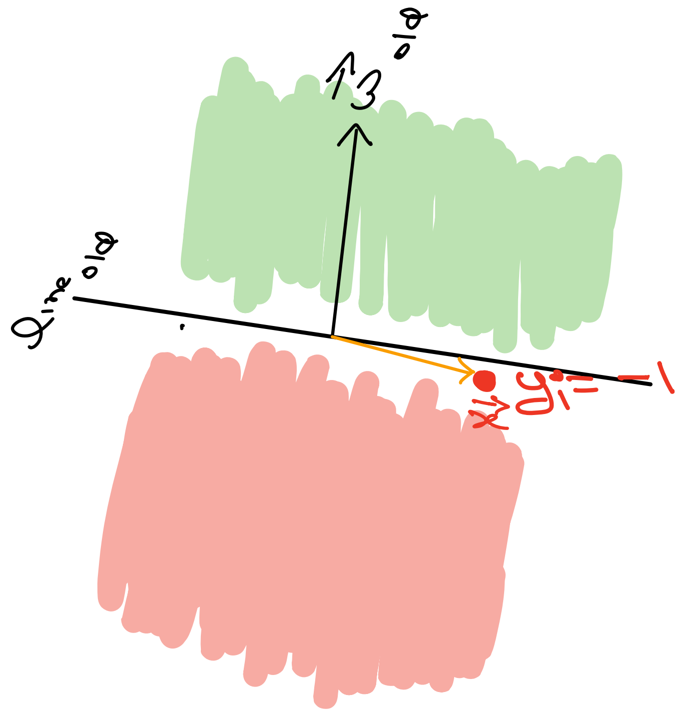
    print ("Correct")

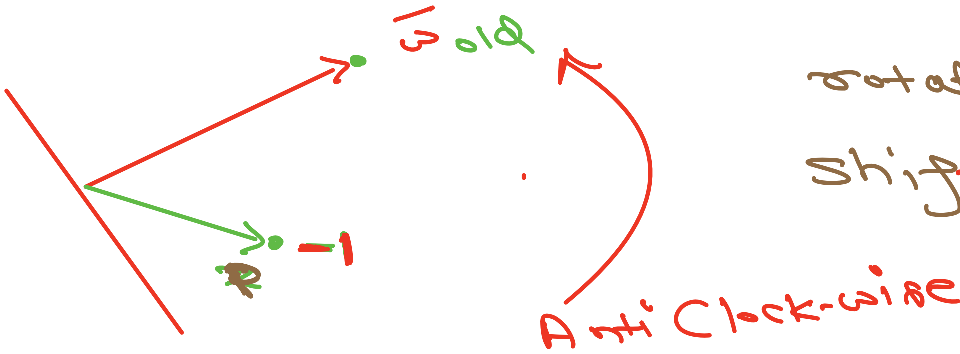
else

    print ("Incorrect")

Q2: How do we reduce misclassification?

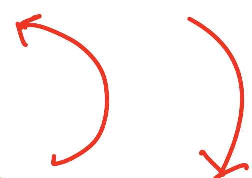






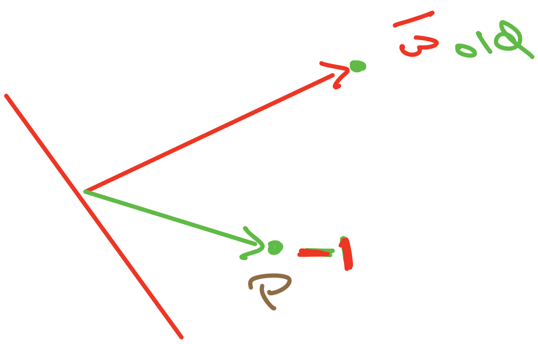
rotation → $\bar{w}$

shift → w_0

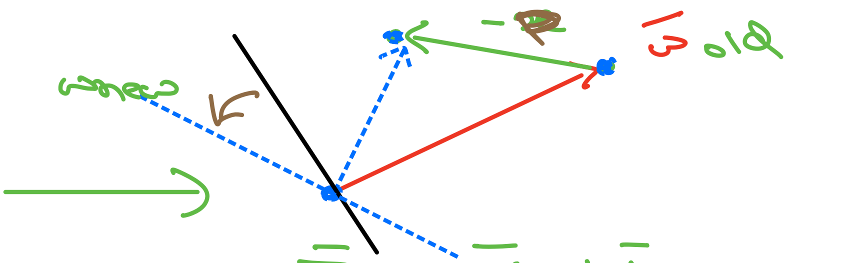


$\bar{w}_{new} = \bar{w}_{old} - \bar{x}$ (Anti-clockwise)

(we are subtracting -ve mis-class)



wold − P



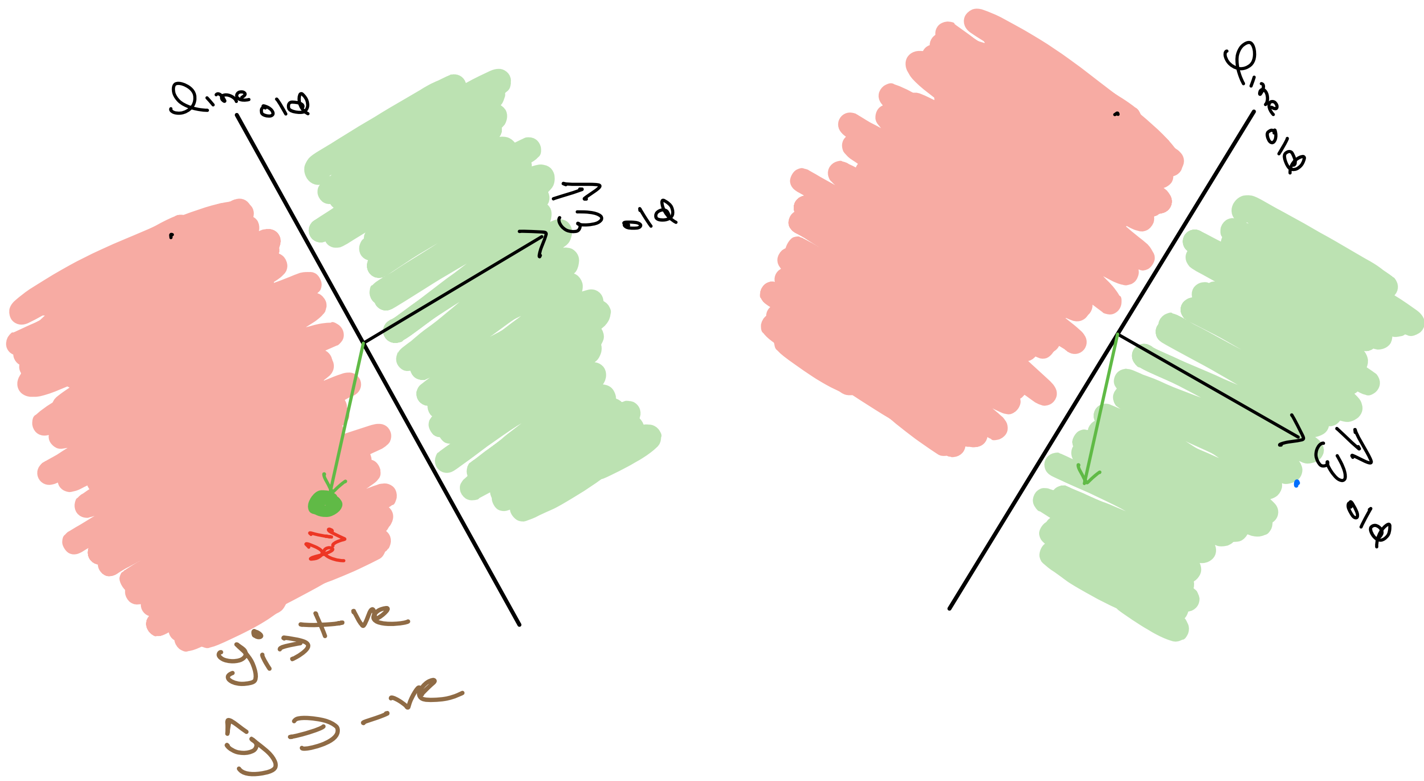
$\bar{z} = \bar{x} + \bar{y}$

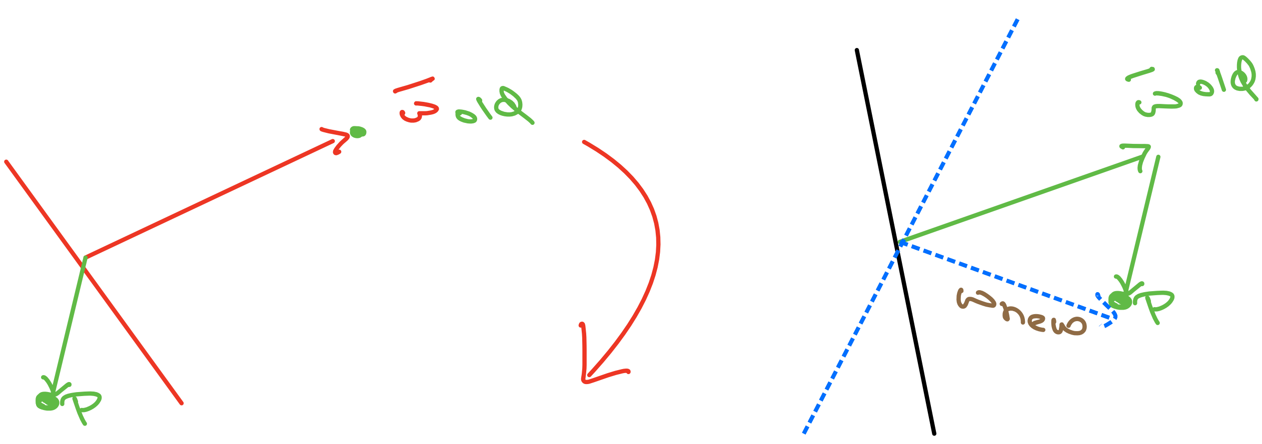
$\bar{w}$_new = $\bar{w}_{old}$ + (−p)

if $y_i = -1$ and $\hat{y} = +1$

$\Rightarrow \quad \bar{w}_{new} = \bar{w}_{old} - \bar{x}$

Case - 2





$w_{new} = w_{old} + \bar{x}$

if $y_i = +1$ and $\hat{y} = -1$

$\Rightarrow \bar{w}_{new} = \bar{w}_{old} + \bar{x}$

x_i and y_i

Combining both of steps into 1 :

Update Step $\Rightarrow$ $\boxed{\bar{w}_{new} \Rightarrow \bar{w}_{old} + \bar{x} \times y_i}$

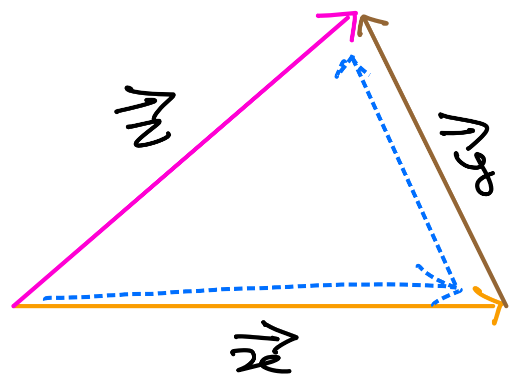
Perceptron

① Initialize w and w_0

② Update using

Vector addition

$$\vec{z} = \vec{x} + \vec{y}$$

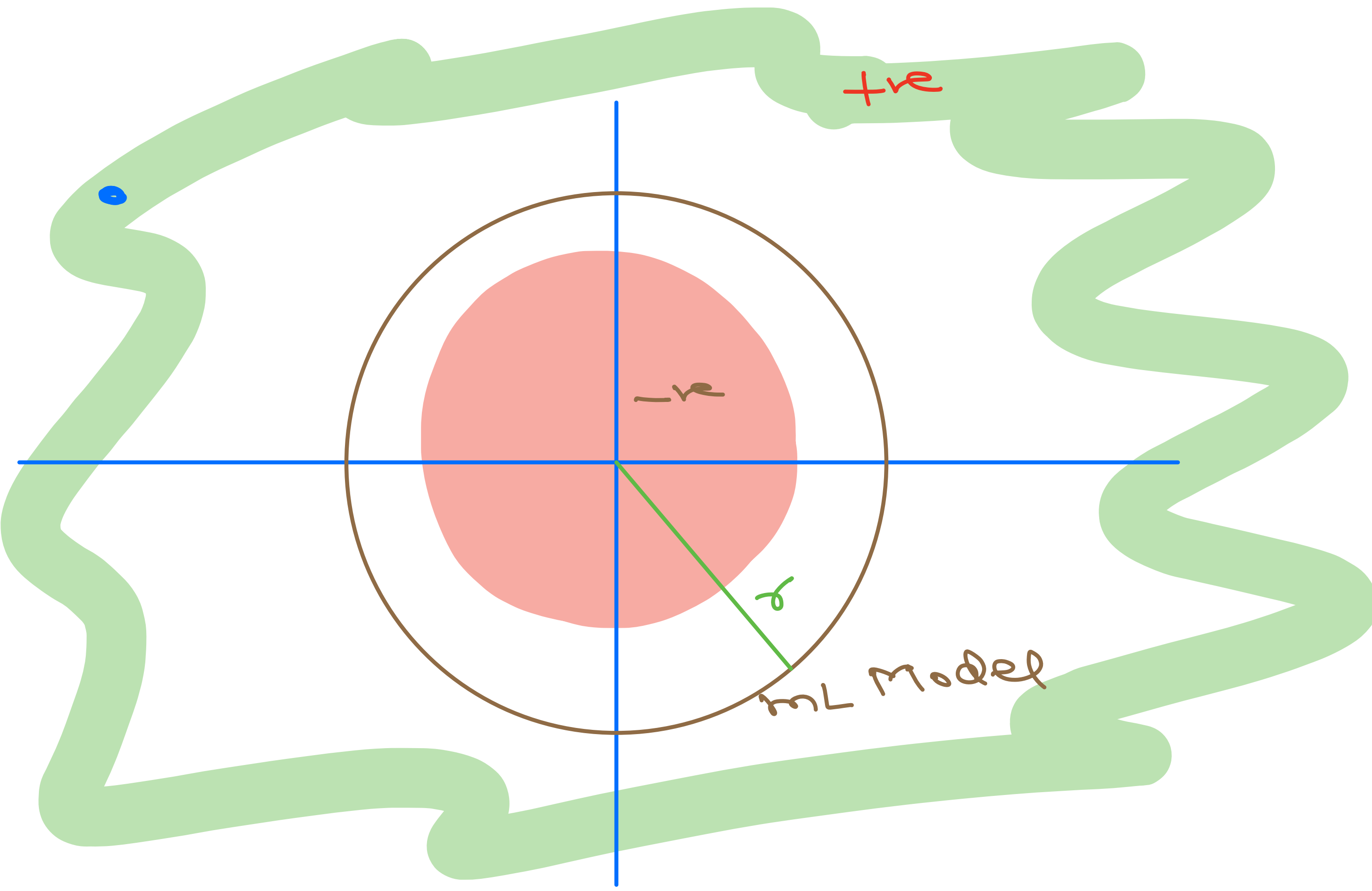


Q Eqn of line ⇒ $3x_1 + 4x_2 + 1 = 0$

feature Vector = X ⇒

x_1	x_2
1	4
-2	-1
3	-1

Which HF do these points lie in?



$\underline{\underline{10}}$ $x^2 + y^2 = r^2$ st $r = 3$

$\hat{y}$:

$\bar{x}$ $-2, 1 \longrightarrow$

$\bar{x}$ $3, 1 \longrightarrow$