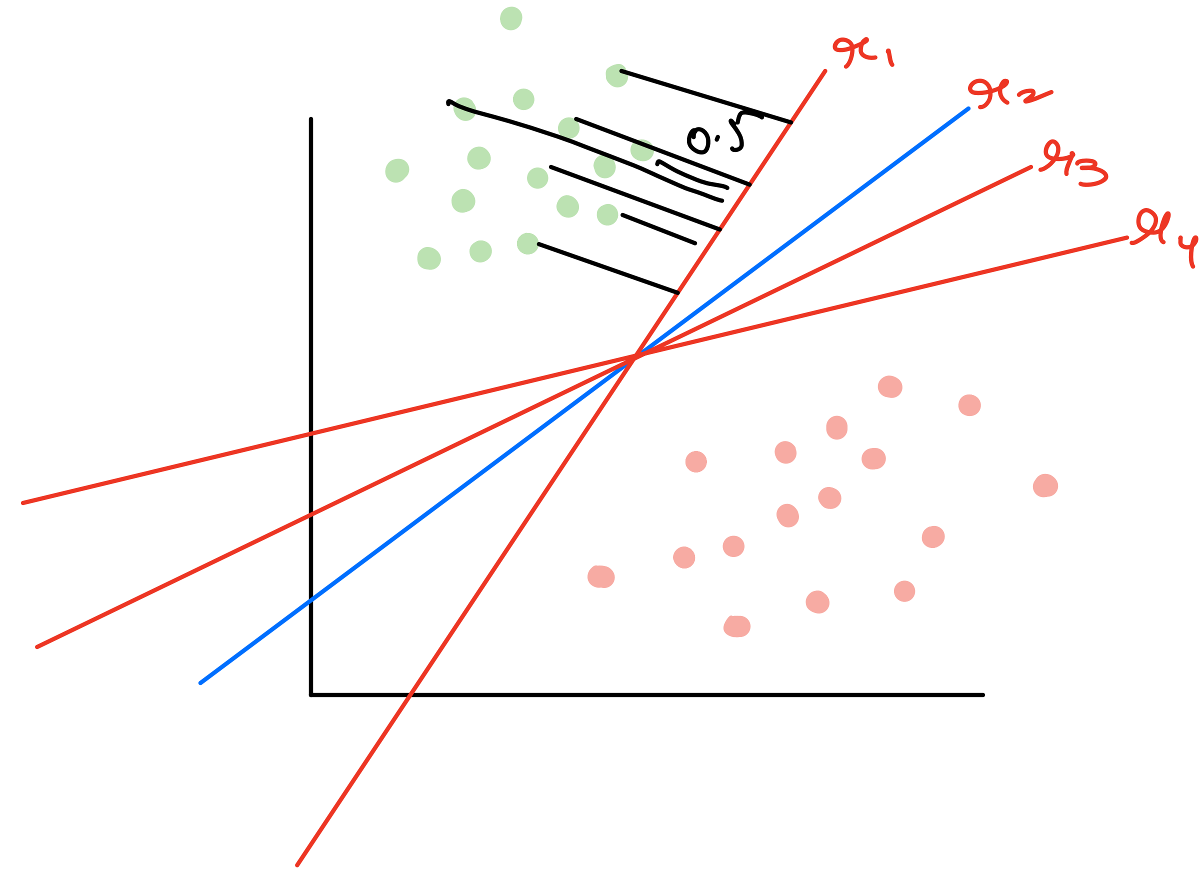
- Agenda
- Maximum Margin Classifier : Geometric Intuition
- SVM : Support Vector Machines
    - Hard Margin Classifier
    - Soft Margin Classifier
- Hinge Loss
- Comparing with Log-Loss
- Effect of Data Imbalance

SVM

- Popular in 2000's
- Good foundation of Maths
- Can be used for Linear as well as Non-Linear Classification

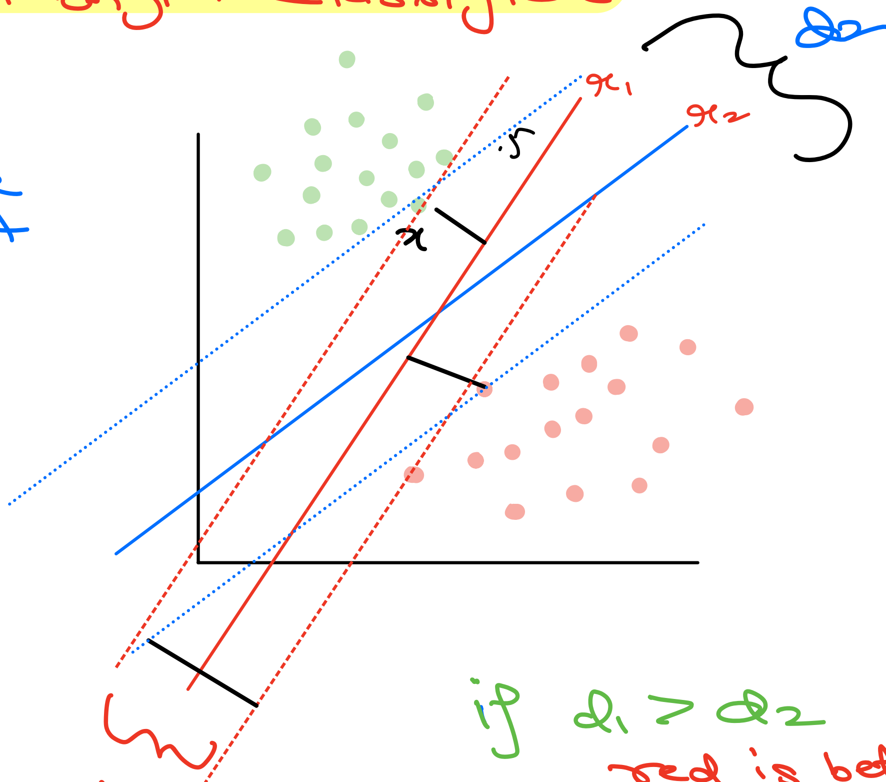
Maximum Margin Classifier



Maximal Margin Classifier

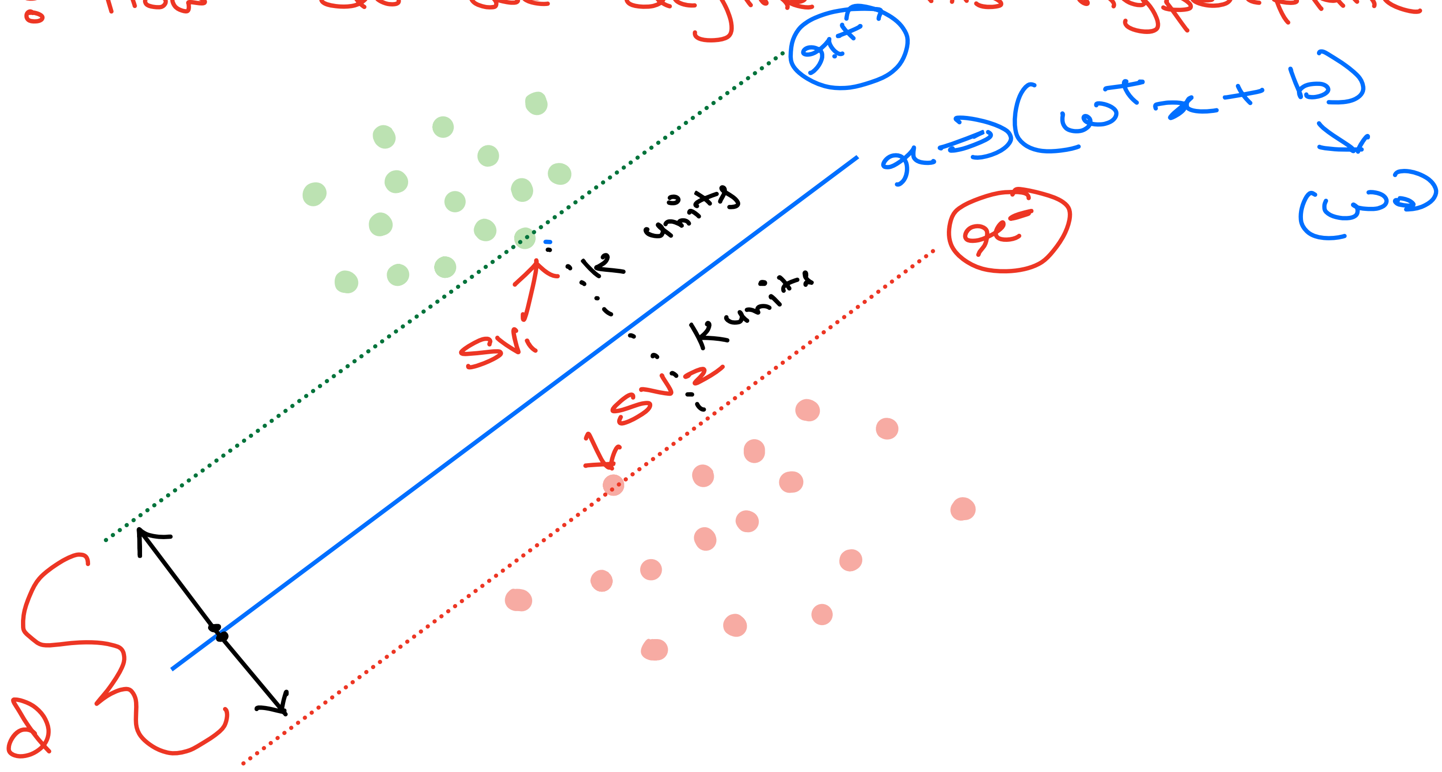
Goal:

1. Find the line with highest Margin
2. Higher the Margin Better the Seperation Between



if $d_1 > d_2$
red is better
else
Blue is better

Q: How do we define this Hyperplane



* The points Closest to the Hyperplane are Called support Vectors

Let, $\pi \Rightarrow w^T x + b = 0$

$\pi^+ \Rightarrow w^T x + b - k = 0 \Rightarrow w^T x + b = k$

$\pi^- \Rightarrow w^T x + b + k = 0 \Rightarrow w^T x + b = -k$

Objective ⇒ Increase Margin

Q: What will be length of Margin?

① Distance of Both x^+ and x^- from origin

$d \to 0$

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

$x^+ : d(0, x^+) \Rightarrow \frac{b-k}{\|w\|}$

$x^- : d(0, x^-) \Rightarrow \frac{b+k}{\|w\|}$

$|\text{Margin}| \Rightarrow d(x^+, x^-)$

$\Rightarrow d(0, x^+) - d(0, x^-)$

$\boxed{d \Rightarrow \frac{2k}{\|w\|}}$ ← maximize

Q: What are parameters of Margin

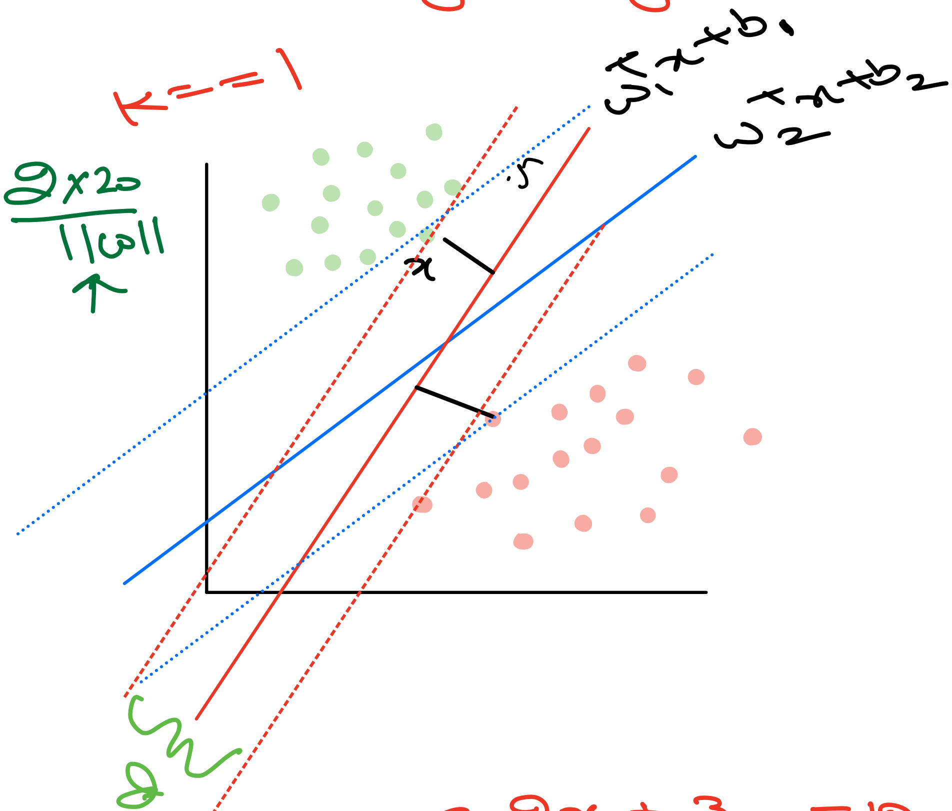
To maximize $K = 1$

$\frac{2}{\|w\|}$ $\frac{2k}{\|w\|} \leftarrow \frac{2 \times 2}{\|w\|}$

Parameters to get maximum Margin

⇒ w

⇒ b



$\begin{cases} 2x + 3y = 10 \\ 4x + 6y \Rightarrow 20 \end{cases}$

$2x$

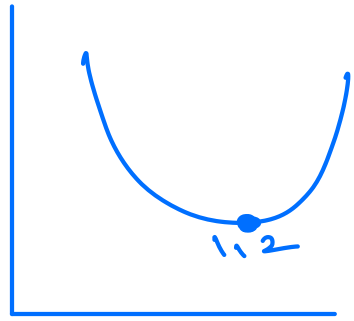
obj $\arg\max_{w,b} \frac{(2k)}{||w||}$

$k \geq 1$

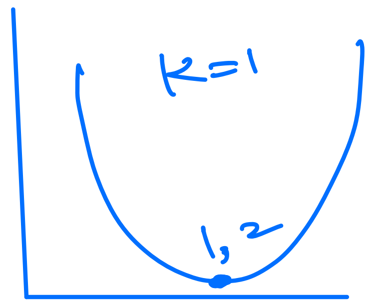
$\arg\max_{w,b} \frac{2}{||w||}$

$k \geq 10$

$\arg\max_{w,b} \frac{20}{||w||}$



Let fix the $k \geq 1$
since k just change the
scale of Margin



Objective

Maximum Margin

$$\arg\max_{w,b} \frac{2}{||w||}$$

↓

$$\boxed{\arg\min_{w,b} \frac{||w||}{2}}$$ ← Loss

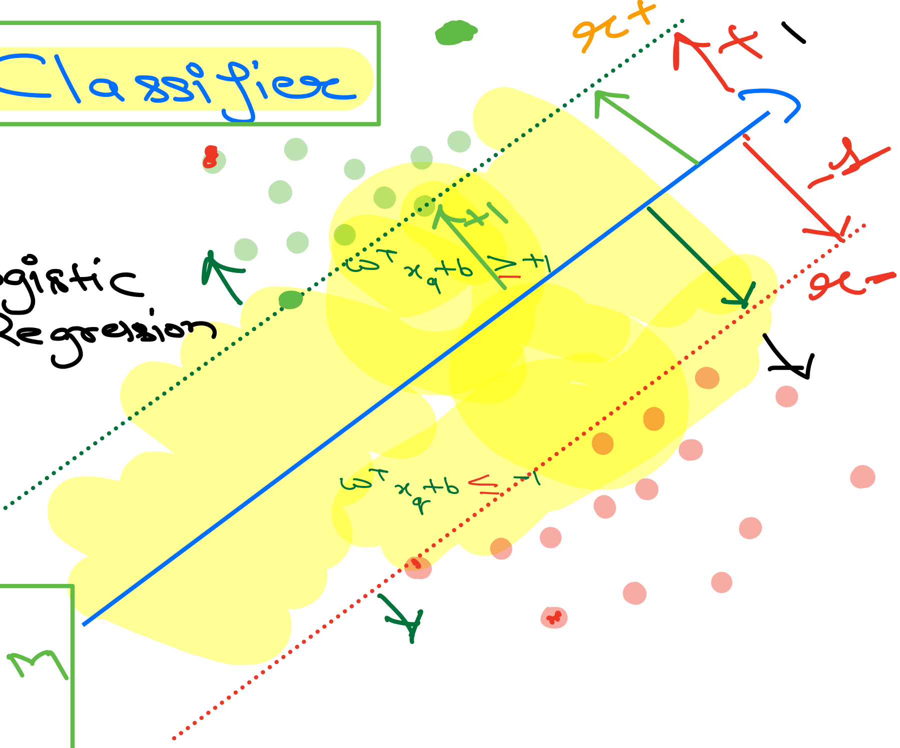
Gain $\frac{2}{100 \times a}$

-1

$\frac{100 \times a}{2}$

Hard Margin Classifier

$y_i \Rightarrow +1$
$y_i \Rightarrow 0$
} Logistic Regression



Classes

$\boxed{\begin{array}{l} y_i \Rightarrow +1 \\ y_i \Rightarrow -1 \end{array} \Bigg\} \text{SVM}}$

Function Margin ⇒ Margin × Label

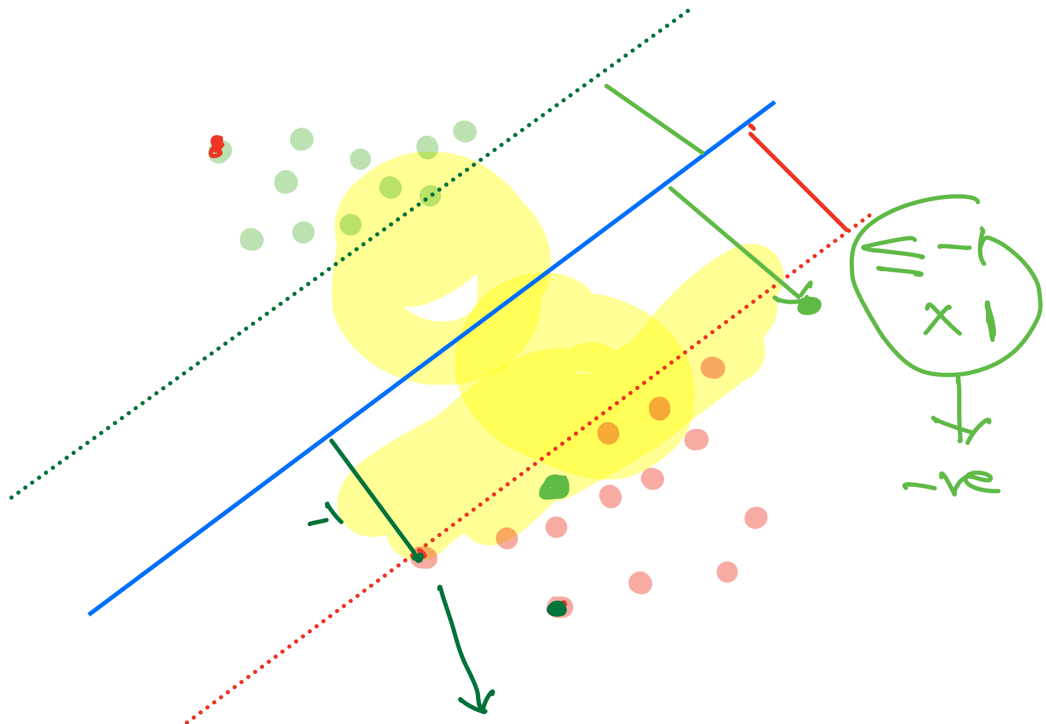
Constraint $x^+ \Rightarrow y_i \Rightarrow (w^T x + b) \geq 1$

$x^- \Rightarrow y_i \Rightarrow (w^T x + b) \leq -1$

* To check for misclassification of a point, we can multiply the point plugged into equation and check for positive

for every point, It should Satisfy

$$\boxed{(w^T x_q + b) \times y_i \geq 1}$$



margin 2 ⇒ $w^T x_q + b \Rightarrow -1$

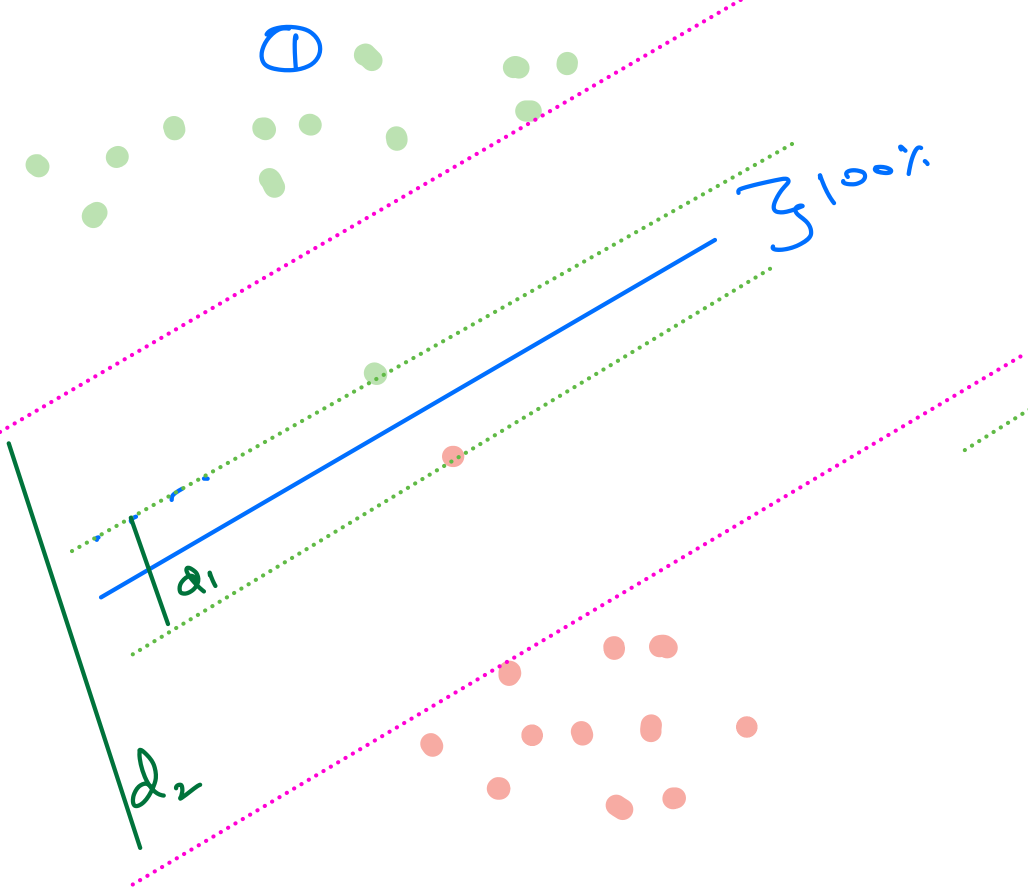
F. margin 2, × y_i ⇒ -1×-1 ⇒ 1

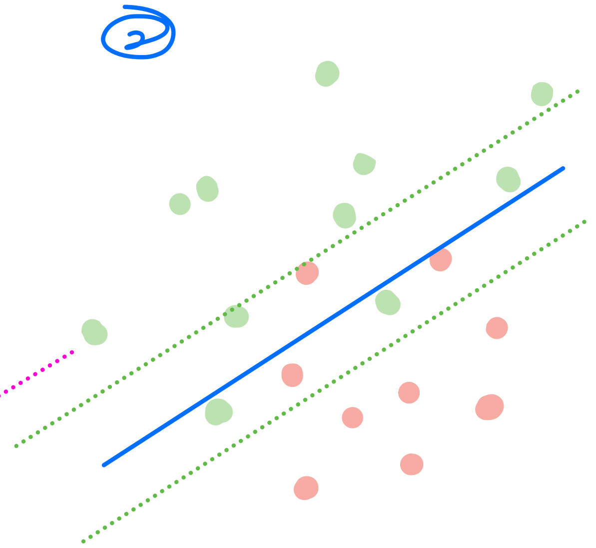
Constrained Optimization

$$\boxed{\underset{}{\operatorname{argmax}}\left(\frac{2}{\|w\|}\right) \quad \text{s.t.} \ \forall\, i : 1-N \quad y_i \times w^T x_i + b \geq 1}$$

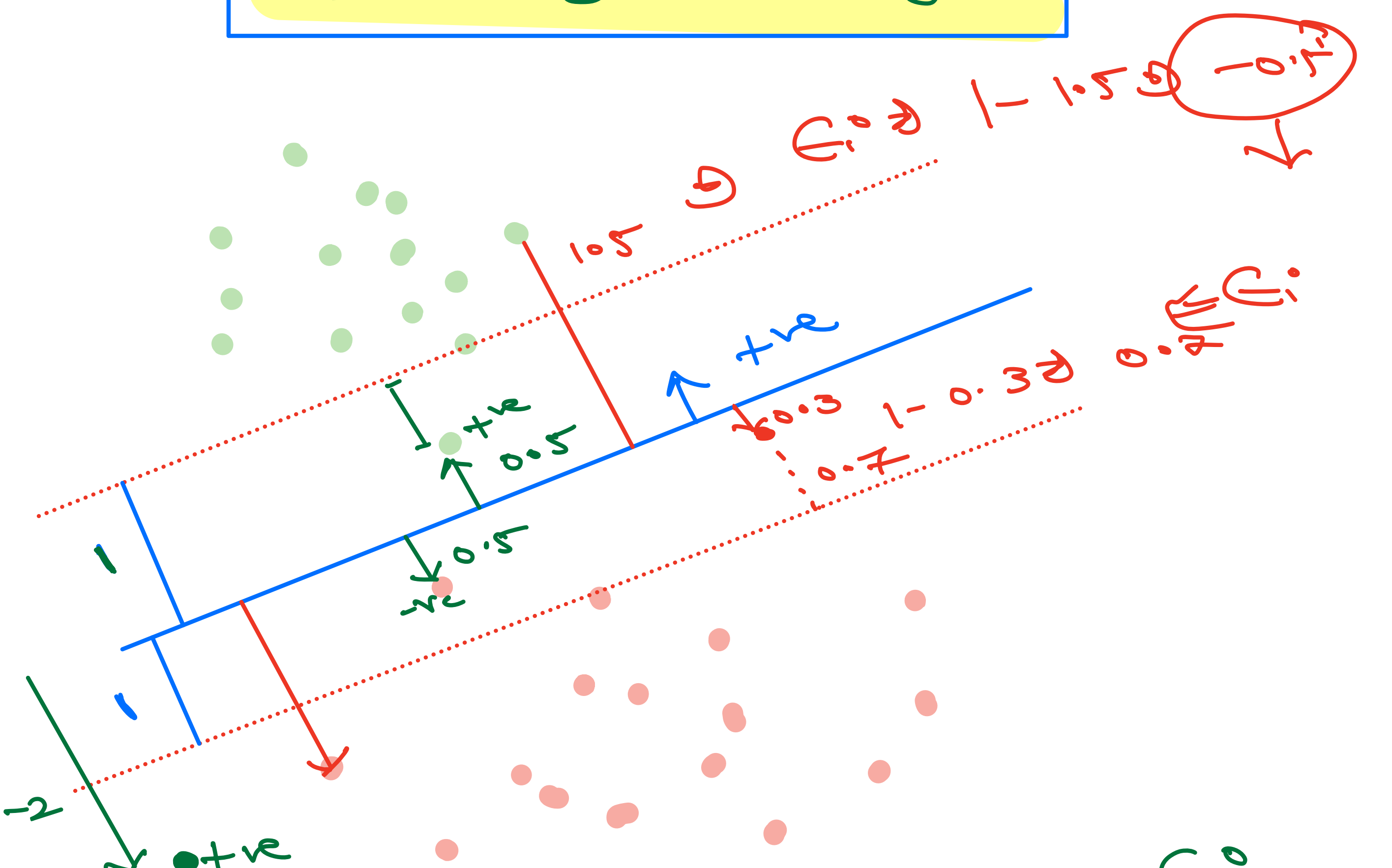
Hard Margin Classifier

It won't allow any point between margins as well as misclassification





Soft Margin Classifier



① $z_i \Rightarrow w^T x_q + b \Rightarrow -2 \times 1 \Rightarrow z_i \Rightarrow -2$

ϵ_i

$1 - (-2) \Rightarrow 3$

$+ve \Rightarrow w^T x_{+ve} + b \Rightarrow 0.5 \Rightarrow z_i$

$\text{Error} \Rightarrow 1 - z_i$

$-ve \Rightarrow w^T x_{-ve} + b \Rightarrow 0.5$

$$\epsilon_i \Rightarrow 1 - \underbrace{z_i}_{w^T x_i + b} \Rightarrow z_i \Rightarrow 1 - \epsilon_i$$

H.M $\Rightarrow$ $z_i \times y_i \geq 1$

Soft margin $\Rightarrow$ $z_i \times y_i \geq 1 - \epsilon_i$

$$(w^T x_i + b) \times y_i \geq 1 - \epsilon_i$$

$$\boxed{\operatorname{argmax}\left(\frac{2}{\|w\|}\right) \quad \text{S.t } \forall\, i : 1 - N \quad y_i \times w^T x_i + b \geq 1 - \epsilon_i \quad \text{and } \epsilon_i \geq 0}$$

Since this allows Errors its call Soft Margi

Hyperparameter

$$\arg\min \frac{\|w\|}{2} + \frac{C}{N} \times \sum_{i=1}^{N} \epsilon_i$$

s.t. $\forall\, i : 1 - N$

$y_i \times w^T x_i + b \geq 1 - \epsilon_i$

and $\epsilon_i \geq 0$

if c is very Low

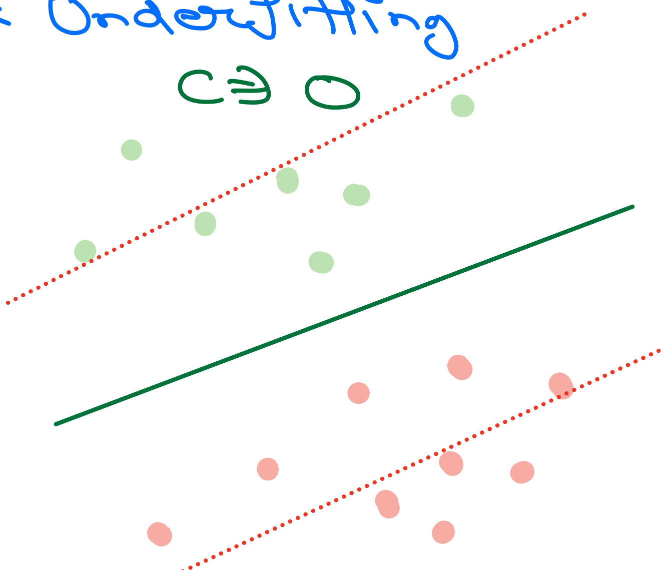
$C \to 0$

1) $\arg\min \frac{\|w\|}{2}$

2) Maximize margin and allow $(1 - \epsilon_i)$

* Underfitting

$C \to 0$



if c is very high

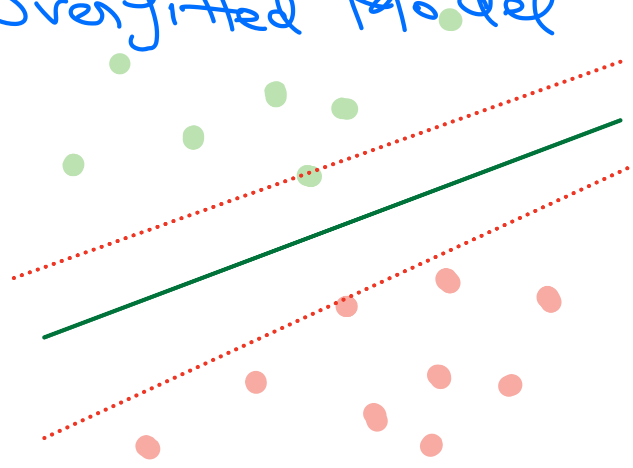
It will become too soft

1) Focuses on

$C \times \sum_{i=1}^{n} \epsilon_i$

2) amount of Errors allowed

3) Overfitted Model



* Hence c should be tuned properly

Soft Margin Classifier

$$\text{argmin}\left(\frac{\|w\|}{2} + - \frac{C}{N} \times \sum_{i=1}^{N} \epsilon_i\right)$$

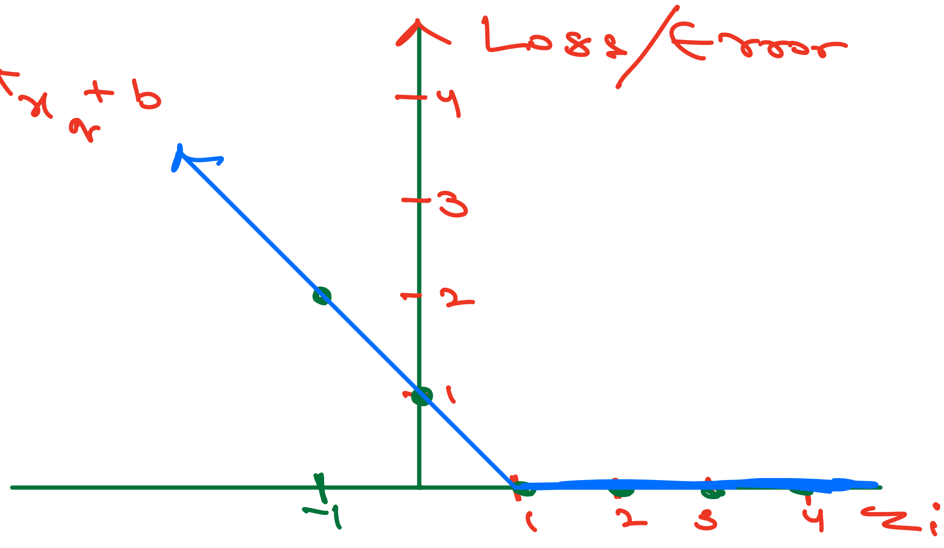
$$\frac{1}{C}\ \frac{\text{regularization}}{\times\ 2} + \text{Loss function}$$

$\Rightarrow$ Soft Margin Classifier

$$\frac{1}{2} L2 + \text{Hinge Loss}$$

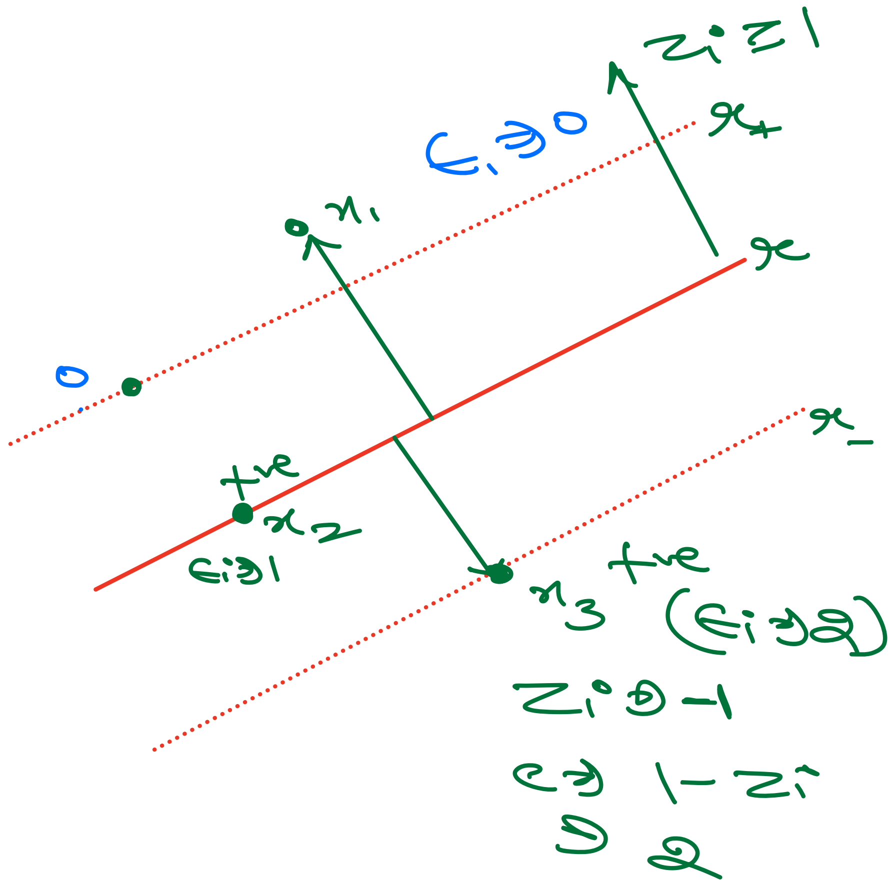
* Intuition of Hinge Loss

$z_i \Rightarrow y_i \times w^T x_i + b$



$1 - \xi_i$

$\xi_i \geq 1 - z_i$



Case 1

- Points x_i that lies beyond x^+
- $y_i(w^T x_i + b) \geq 1$
- $\xi_i \geq 0$

Case 2

- Points that lie on x
- $y \times w^T x + b \geq 0$
- $\xi_i \geq 1$

Case 3

- Points that lie on the other side of π
- $y \times w^T x + b < 0$
- $\xi_i \geq$ +ve int

Hinge Loss

$$\boxed{\xi_i \geq \max(0, 1 - z_i)} \leftarrow \xi_i \geq 0$$

Comparison with Logistic Regression

Logistic	S.V.M
① $y \in (0, 1)$	① $y \in (-1, 1)$
② Log Loss + Reg	② Hinge Loss + Reg
$\sum_{i=1}^{n} y_i \times \log \hat{y} + (1-y_i) \times \log(1-\hat{y})$	$\sum_{i=1}^{n} \max(0, 1 - Z_i)$
+	+
$\lambda \lVert w \rVert$	$\frac{1}{\lambda 2} \times \lVert w \rVert$

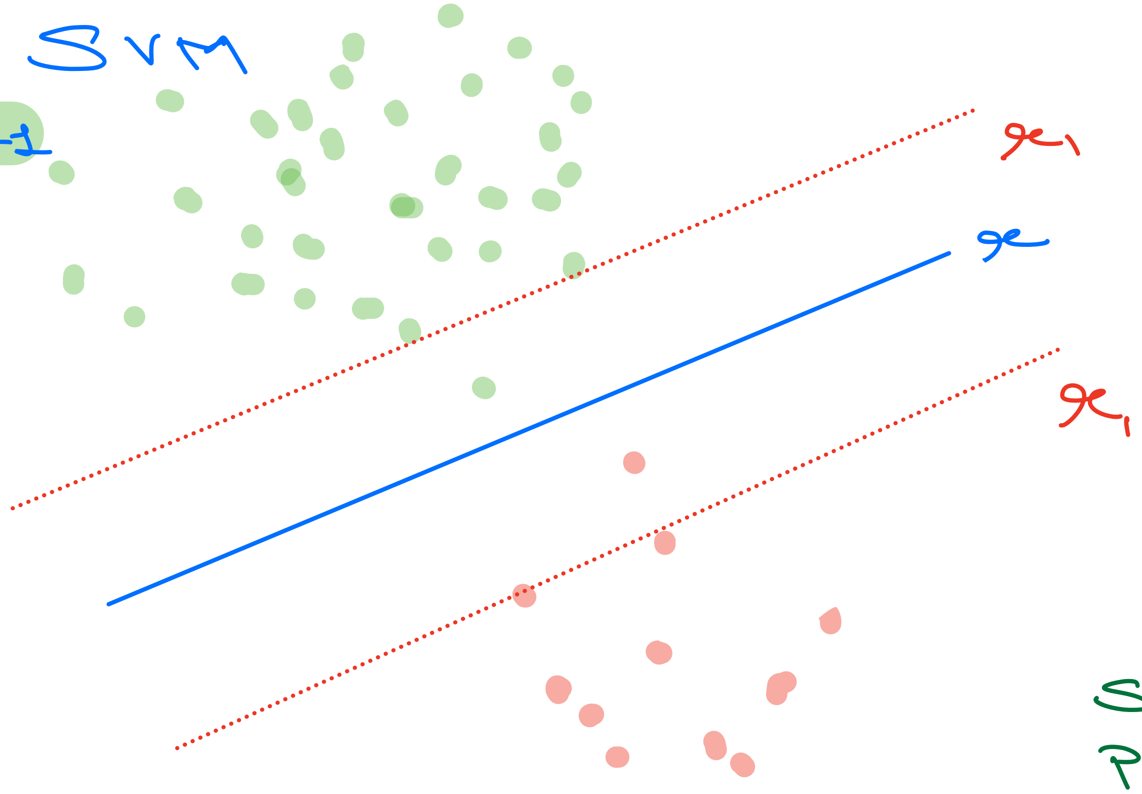
Q) Implement SVM on Spam vs Ham Dataset

```
From sklearn.svc import SVC
```

1) Tune the C parameter to find best value between 0 and 10

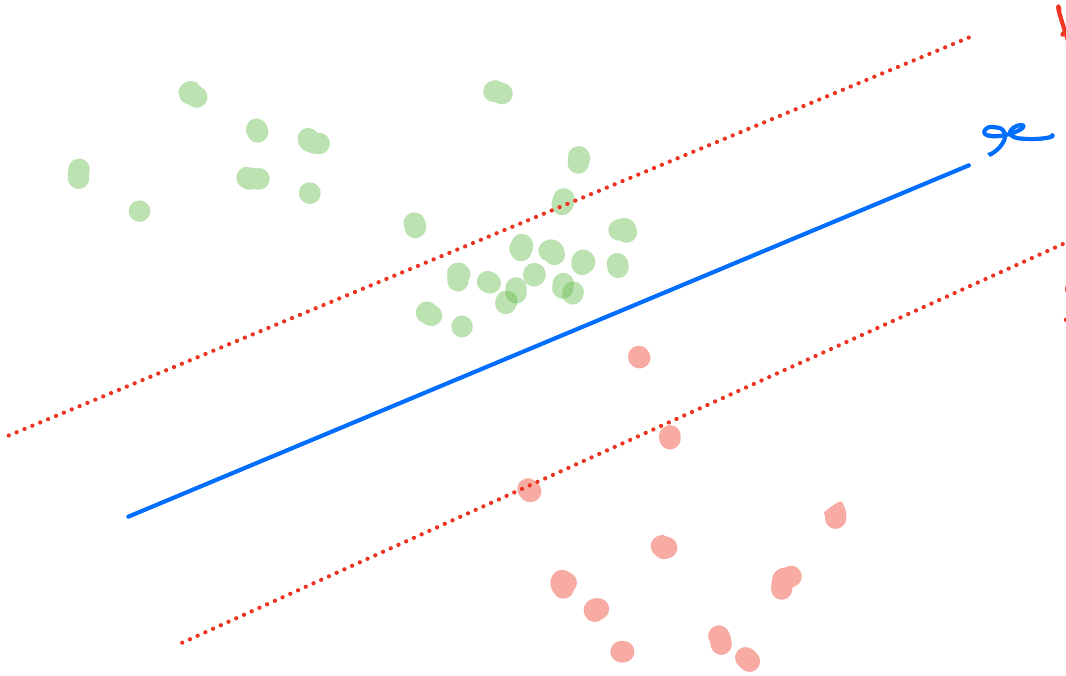
Impact of Data Imbalance on SVM

Case-1



S.V points → 6

Case 2



High C → Less Misclassification

* SVM is impacted by Data imbalance in support Vectors Only.

Topics for Next Session

1) Dual Form

2) Non Linear SVM via Kernels