05 SVM - Pegasos Algorithms For Unconstrained optimization :

- Non - Convex

  ↓

  Convex with constraints

  ↓

  Convex with Constraints + Outliers

Primal formulation

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

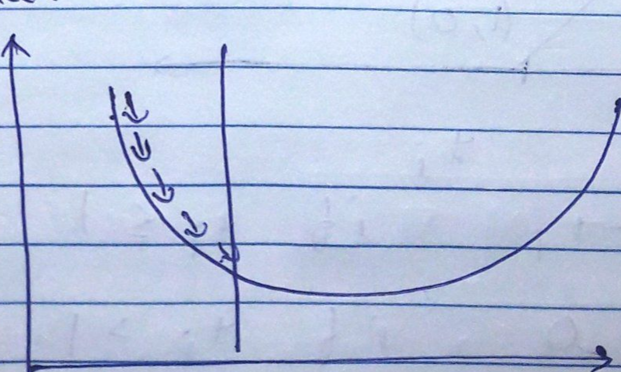
$\Rightarrow$ Lagrangian Duality : Dual formulation

- New Paper in 2011

Pegasos as Unconstrained Convex Optimization

Pegasos : Primal
Estimated
sub-GrAdient
SOlver for
Support Vector Machines

subGradient :



Normal Gradient Descent can't be applied due to Constraint.

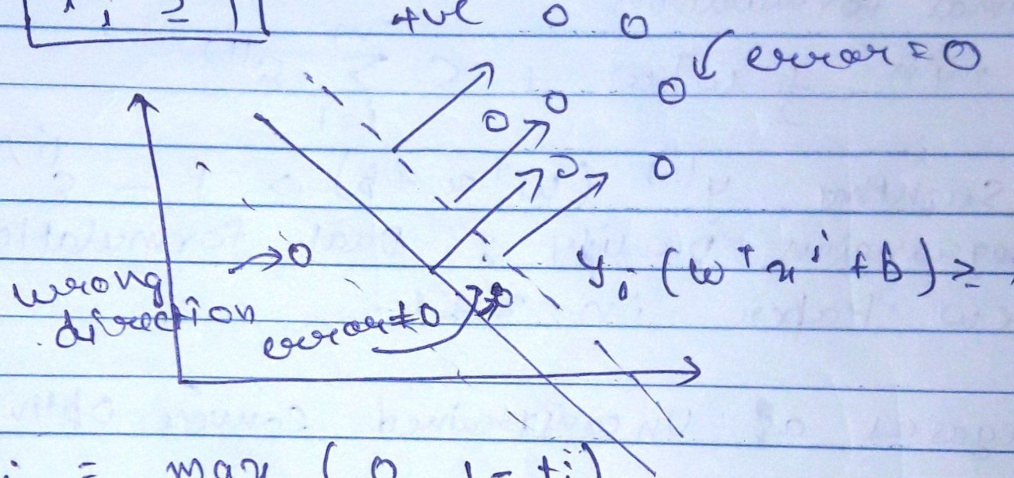
error

① $\varepsilon^{(i)} \geq 1 - \underbrace{y^{(i)}\left(w^T x^{(i)} + b\right)}_{t_i \text{ (suppose)}}$

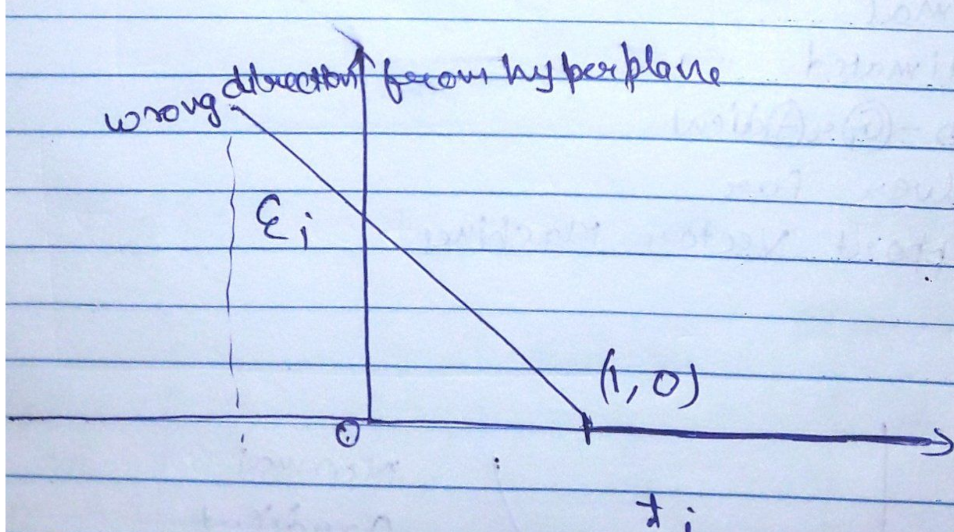
② $\varepsilon^{(i)} \geq 0$

$\varepsilon_i \geq 1 - t_i$

- if $\boxed{t_i \geq 1}$ +ve



$\varepsilon_i = \max(0, 1 - t_i)$



$\varepsilon_i = 1 - t_i \quad \text{if } t_i \leq 1$

$\varepsilon_i = 0 \quad \text{if } t_i > 1$

$$\xi_i = \max(0, 1 - t_i)$$

Loss / SVM objective =

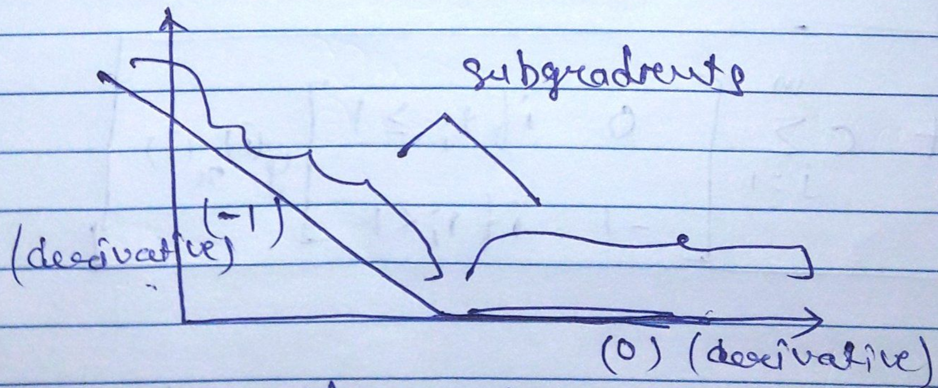
$$\Rightarrow \left[\begin{array}{l} \min_{w,b} \; \frac{1}{2} w^T w + C \sum_{i=1}^{m} \max(0, 1 - t_i) \\ \\ \text{where } \; t_i = y^{(i)} (w^T x^{(i)} + b) . \end{array} \right]$$

This is now unconstrained convex optimisation problem.
Now we can apply gradient descent directly

$w, b = ?$ (parameters)

$w = \text{init}()$

$$\left[w = w - \eta \nabla_w L \right] \leftarrow \text{Gradient Descent}$$



Not differentiable at $t_i = 1$

So we will differentiate (derivative) in two parts.

$$\nabla_\omega L = \omega + C \sum_{j=1}^{m} \nabla_\omega \left[\max(0, 1 - t_i)\right]$$

$$= \omega + C \sum_{i=1}^{m} \frac{\partial f}{\partial t_i} \cdot \frac{\partial t_i}{\partial \omega} \quad \text{(chainrule)}$$

$$= \omega + C \sum_{j=1}^{m} \nabla_t \left[(\max(0, 1 - t_i))\right] \nabla_\omega t_i$$

$$= \omega + C \sum_{j=1}^{m} \begin{bmatrix} 0 & \text{if } t_i \geq 1 \\ -1 & \text{if } t_i < 1 \end{bmatrix} y^{(i)} x^{(i)}$$

$$\left[\begin{aligned} t_i &= y_i (w^T x^i + b) \\ \nabla_w t_i &= y_i x^i \end{aligned}\right]$$

$$\boxed{\nabla_\omega L = \omega + C \sum_{j=1}^{m} \begin{bmatrix} 0 & \text{if } t_i \geq 1 \\ -1 & \text{if } t_i < 1 \end{bmatrix} y^{(i)} x^{(i)}}$$

putting in

$$\boxed{\omega = \omega - \eta \nabla_\omega L} \quad \text{Gradient descent}$$

This type of loss in known as hinge loss