Solving Optimization problem for SVM

Problem P_0

$$\min_{w,b} \frac{\|w\|^2}{2} \longrightarrow \text{convex}$$

$$\text{s.t.} \quad y_i(w^T x_i + b) \geq 1 \longrightarrow \text{linear} \qquad \text{—— (1)}$$

We represent it as a Lagrangian primal problem.

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \lambda_i \left(y_i(w^T x_i + b) - 1 \right) \quad \{\lambda_i \geq 0\} \qquad \text{—— (2)}$$

λ_i → Lagrangian multipliers

To minimize L_p with respect to w, b

$$\frac{\partial L_p}{\partial w} = 0 \implies w - \sum_{i=1}^{n} \lambda_i y_i x_i = 0 \qquad \text{—— (3)}$$

$$\frac{\partial L_p}{\partial b} = 0 \implies \sum_{i=1}^{n} \lambda_i y_i = 0 \qquad \text{—— (4)}$$

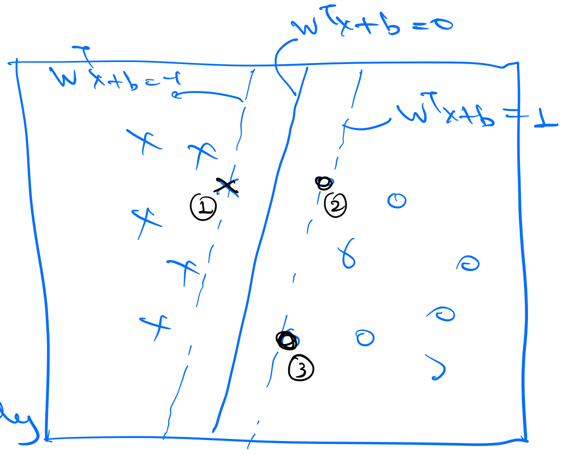
from Karush–Kuhn–Tucker Conditions we get → for better problem conditioning

$$\lambda_i \left[y_i(w^T x_i + b) - 1 \right] = 0 \qquad \text{—— (5)}$$

$\Rightarrow \lambda_i = 0$ if $y_i(w^T x_i + b) \neq 1$

i.e except sample 1, 2, 3 for all other samples $\lambda_i = 0$

Sample 1, 2, 3 referred to as Support vectors. & exclusively control the separation boundary



X : $y_i = -1$

o : $y_i = +1$

Hence, we need to find λ_i for all datapoints to find which data points are contributing to the boundary.

$\Longrightarrow$ we solve an equivalent problem (dual) which is just dependent of λ_i's

$$L_d = \frac{1}{2} \sum_{i=1}^{n} \lambda_i y_i x_i \cdot \sum_{j}^{n} \lambda_j y_j x_j$$

$$- \sum_{i=1}^{n} \lambda_i \left(y_i \left(\sum_{j=1}^{n} \lambda_j y_j x_j \cdot x_i + b \right) - 1 \right)$$

$$L_d = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j \langle x_i \cdot x_j \rangle$$

$$- \sum_{i=1}^{n} \lambda_i y_i \left(\sum_{j=1}^{n} \lambda_j y_j x_j \right) x_i - \cancel{\sum_{i=1}^{n} \lambda_i y_i b}^{\,0} \quad \left\{ \text{from } 4 \right\}$$

$$+ \sum_{i=1}^{n} \lambda_i$$

$$L_d = \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle$$

Problem P_1

$\max \; L_d$

Subject to $\sum_{i=1}^{n} \lambda_i y_i = 0$

$\lambda_i \geq 0$ $\Rightarrow$ quadratic programming problem

$$\left\{ \begin{array}{c} \text{min Primal problem} \\ \Longleftrightarrow \\ \text{max dual problem} \end{array} \right\}$$

1) solve problem P_1 to find λ_i & substitute in (3) to get $\hat{w}$

(2) from (5) & using computed w

$$\hat{b} = \frac{1 - y_i w^T x_i}{y_i}$$

let $S = \{ i \mid \lambda_i > 0 \}$ indices of support vectors

$$\Rightarrow \boxed{\hat{b} = \frac{1}{|S|} \sum_{i \in S} \frac{1 - y_i w^T x_i}{y_i}}$$

$\Rightarrow$ Hence we have our hyperplane as

$$f = \hat{w}^T x + \hat{b} = 0$$

Improvements over basic SVM

1) Soft margin SVM (Better generalization)

$$\min_{w, b, \xi_i} \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \xi_i$$

subject to $y_i (w^T x_i + b) \geq 1 - \xi_i$

$\xi_i \geq 0$

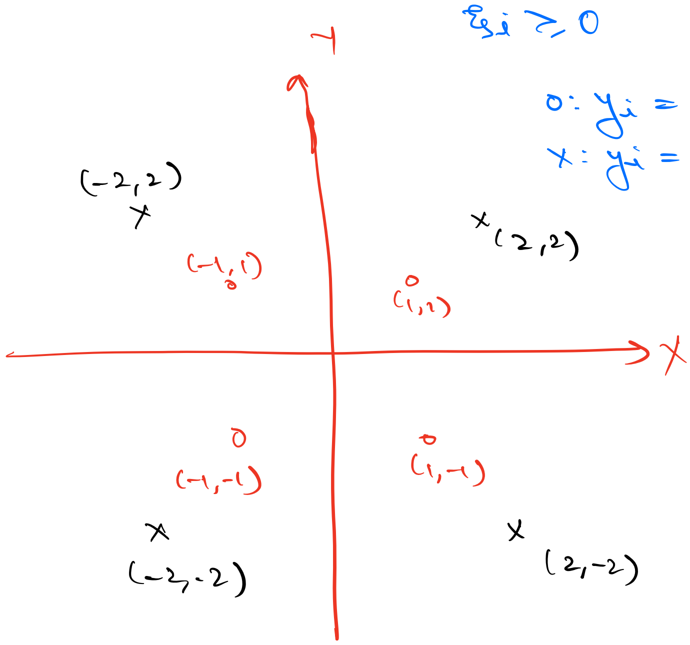
$\rightarrow$ allows some training error

2) Nonlinear SVM (Project data into feature space)

$$\min_{w, b, \xi_i} \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t} \quad y_i \left(w^T \Phi(x_i) + b\right) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$



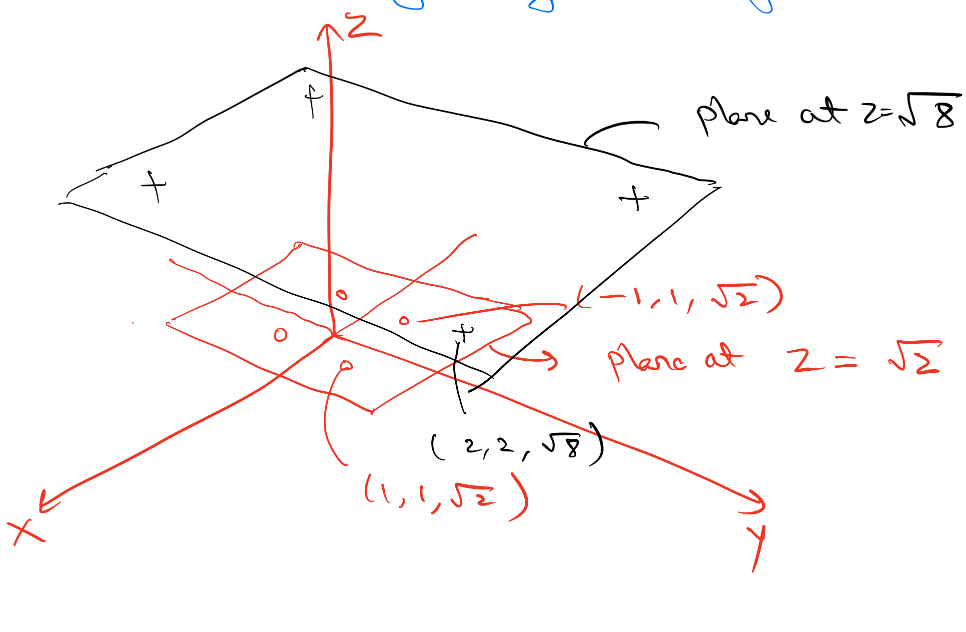
o : $y_i = -1$

x : $y_i = +1$

: No linear hyperplane → can separate the two classes

Make a transformation $\Phi([x, y]) \longrightarrow [x, y, r]$

where r is distance of (x, y) from origin in the original space



any plane perpendicular to z axis & passing between $z = \sqrt{2}$ & $\sqrt{8}$ will perfectly classify the 2 classes.