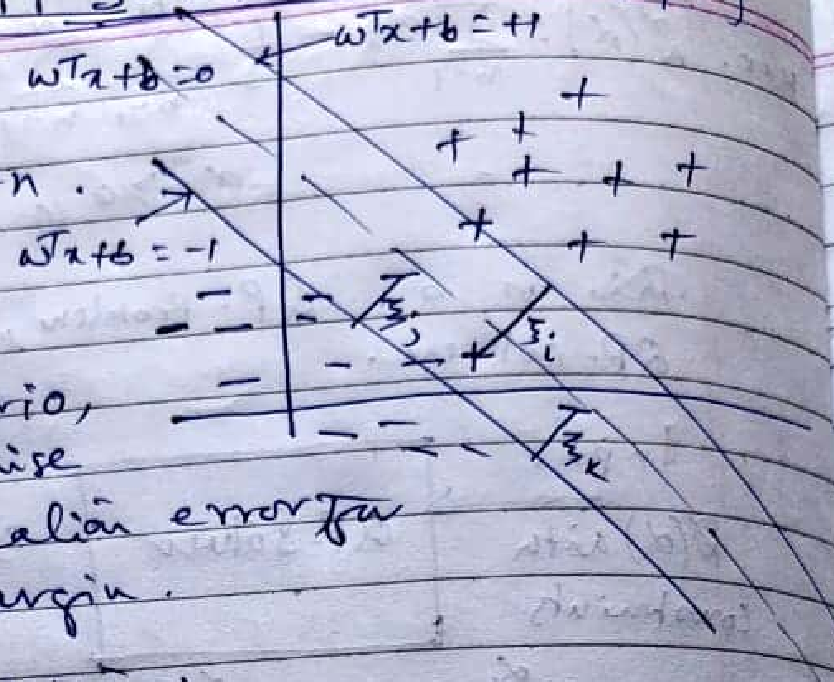
SVM: SOFT MARGIN (1/2)

$\text{Min } \frac{1}{2} w^T w$,

s.t. $y_n(w^T x_n + b) \geq 1, \forall n$.



In some scenario, we compromise a few classification error for a better margin.

The new condition:

$$w^T x_n + b \geq 1 - \xi_n, \quad y_n = +1$$

$$w^T x_n + b \leq -1 + \xi_n, \quad y_n = -1.$$

$$\xi_n \geq 0. \leftarrow \text{slack variables.}$$

Problem becomes:

$$\text{Min } \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n$$

$$\text{s.t. } y_n(w^T x_n + b) \geq 1 - \xi_n, \quad \forall n = 1, 2, \ldots, N$$

$$\xi_n \geq 0,$$

C: Tradeoff parameter between error & margin

$$\mathcal{L} = \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n + \sum_{n=1}^{N} \alpha_n \left(1 - \xi_n - y_n(w^T x_n + b)\right) - \sum_{n=1}^{N} \mu_n \xi_n$$

$\alpha_n \geq 0, \mu_n \geq 0 \; \forall n$ are the Langrange multipliers.

$$\frac{\partial L}{\partial w} = w - \sum_{n=1}^{N} \alpha_n y_n x_n = 0 \Rightarrow w = \sum_{n=1}^{N} \alpha_n y_n x_n .$$

$$\frac{\partial L}{\partial b} = 0 - \sum_{n=1}^{N} \alpha_n y_n = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n y_n = 0 .$$

$$\frac{\partial L}{\partial \xi_n} = C - \alpha_n - \mu_n = 0 \quad \forall n .$$

$$L = \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m \alpha_n y_m y_n x_m^T x_n + C \sum_{n=1}^{N} \xi_n$$

$$+ \sum_{n=1}^{N} \alpha_n \pm \sum_{n=1}^{N} \alpha_n \xi_n - \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m \alpha_n y_m y_n x_m^T x_n$$

$$- b \sum_{n=1}^{N} \alpha_n y_n - \sum_{n=1}^{N} \mu_n \xi_n$$

$$= \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m \alpha_n y_m y_n x_m^T x_n - b \sum_{n=1}^{N} \alpha_n y_n$$

$$+ \sum_{n=1}^{N} (C - \alpha_n - \mu_n) \xi_n$$

$$= \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{m=1}^{N} \sum_{n=1}^{N} \alpha_m \alpha_n y_m y_n x_m^T x_n$$

$$\alpha_n \geq 0, \ \mu_n \geq 0, \quad \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\rightarrow 0 \leq \alpha_n \leq C$$

Use QP-solver to solve for α_n's.

* C is small margin is large & error is more.
* C is large hard margin & no error.

Soft margin is more robust.