COL 774
Machine Learning
Sep 24, 2021

Last class:- Support Vector Machines

$$\min_{w,b} \ \frac{1}{2} w^T w$$

$$\underbrace{y^{(i)}(w^T x^{(i)} + b)}_{\text{functional margin}} \geq 1$$

Assumption: data is linearly separable

↳ Method of Lagrange multipliers:-

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum \alpha_i \alpha_j x^{(i)T} x^{(j)} y^{(i)} y^{(j)}$$

$$0 \leq \alpha_i \quad \forall i \qquad \sum_{i=1}^{m} \alpha_i y^{(i)}$$

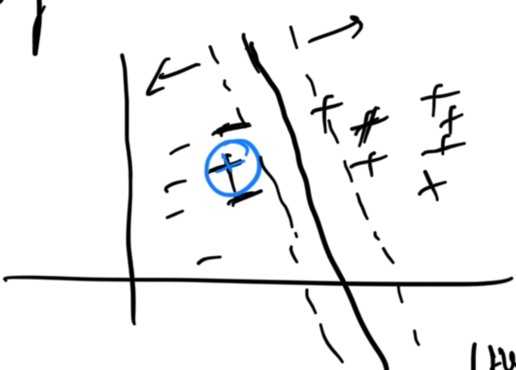
$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

$\cancel{b = 1/2}$

$$b = -\frac{1}{2}\left[\min_{i: y^{(i)}=1} w^T x^{(i)} + \max_{i: y^{(i)}=-1} w^T x^{(i)}\right]$$

Handling noise (in the data)

↳



:- data is not linearly separable

Note:- we are still looking for a linear classifier

Linear SVMs

Hard SVMs

SVMs.

hyperparameter constant

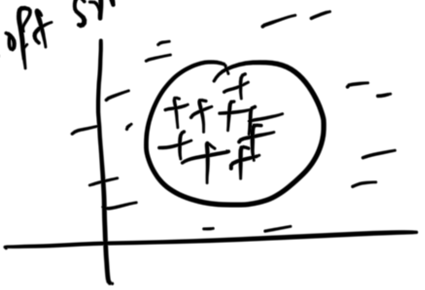
$$\min_{w,b,\xi} \ \frac{1}{2} w^T w + C \cdot \sum_{i=1}^{m} \xi_i$$ penalty / loss

slack

soft SVM $\Big\{$ Hard SVM $\quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \xi_i$:- slack variables

$\xi_i \geq 0$

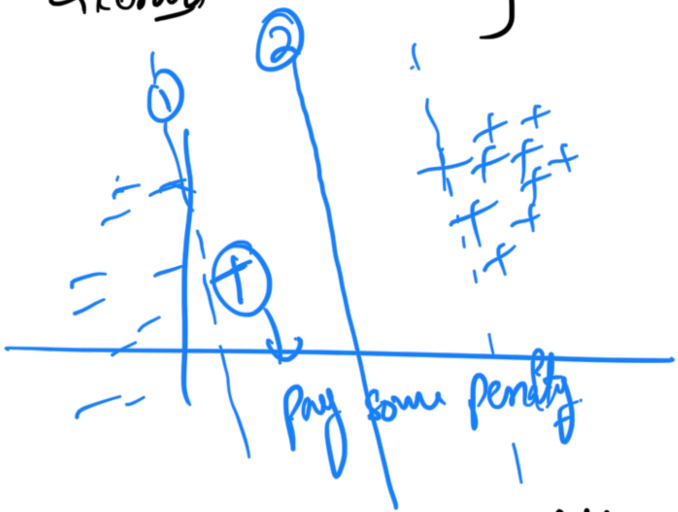
if $C = \infty$ soft SVMs $\rightarrow$ Hard SVMs



Non-linear separation } Later

↳ kernels

Aside:



$$L(w, b, \alpha, \gamma) = \tfrac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \alpha_i \left[1 - y^{(i)} (w^T x^{(i)} + b) - \xi_i \right] + \sum_{i=1}^{m} \gamma_i (-\xi_i)$$

Primal:-

$$\min_{w, b, \xi} \; \max_{\substack{\alpha, \gamma \\ \alpha, \gamma \geq 0}} L(w, b, \alpha, \gamma)$$

Dual

$$\max_{\substack{\alpha, \gamma \\ \alpha, \gamma \geq 0}} \left[\min_{w, b, \xi} L(w, b, \alpha, \gamma) \right]$$

Dual:- $\quad \nabla_w L(w, b, \alpha, \xi) = 0 \qquad \nabla_b L(w, b, \alpha, \xi) = 0$

$\nabla_{\xi} L(w, b, \alpha, \xi) = 0$

Solve:- (HW)

$$w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \qquad \sum_{i=1}^{m} \alpha_i y^{(i)} = 0$$

$$\alpha_i + r_i = C$$ } Additional constraint (vis-a-vis Hard SVMs)

$$\alpha_i \geq 0 \qquad r_i \geq 0$$

$$\Rightarrow \underbrace{0 \leq \alpha_i \leq C} \qquad \underbrace{0 \leq r_i \leq C} \qquad \alpha_i + r_i = C \quad \forall i$$

Complementary Slackness:-

$$\alpha_i \left(\underbrace{1 - y^{(i)} (w^T x^{(i)} + b)} \right) = 0 \qquad \forall i$$

$$r_i (-\xi_i) = 0 \qquad \forall i$$

A $\underbrace{\alpha_i = 0} \Rightarrow r_i = C \Rightarrow \xi_i = 0$ — (1)

$$y^{(i)} (w^T x^{(i)} + b) \geq 1 \qquad w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)}$$

B $\underline{r_i = 0} \Rightarrow \underline{\alpha_i = C} \Rightarrow 1 - y^{(i)} (w^T x^{(i)} + b) - \xi_i = 0$

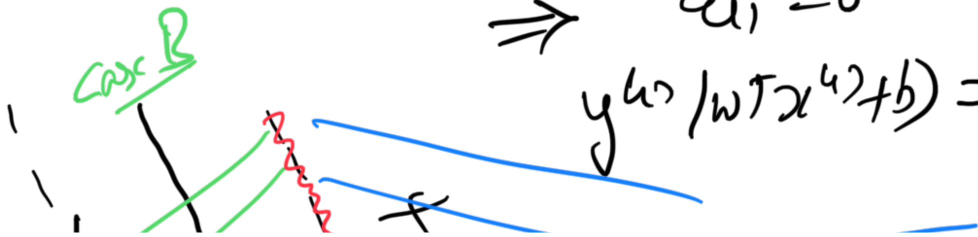
$$\Rightarrow y^{(i)} (w^T x^{(i)} + b) = 1 - \xi_i$$

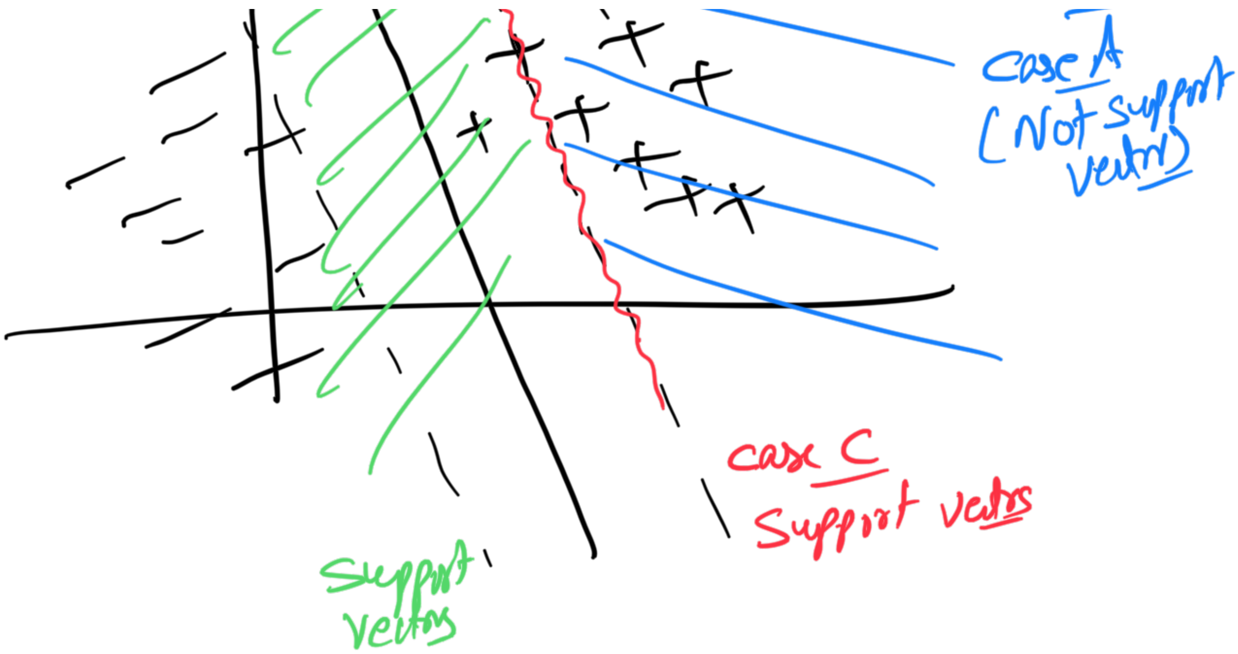
$$\xi_i \geq 0$$

C) $0 < \alpha_i < C \Rightarrow 0 < r_i < C$

$$\Rightarrow \xi_i = 0$$

$$y^{(i)} (w^T x^{(i)} + b) = 1$$





Dual:-

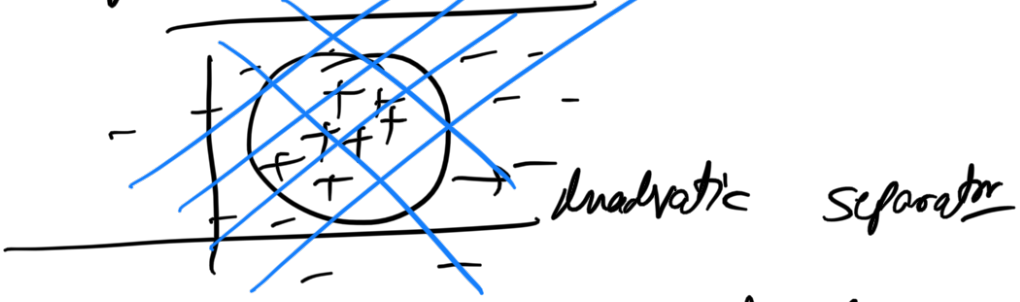
$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)T} x^{(j)}$$

$$\sum_{i=1}^{m} \alpha_i y^{(i)} \qquad 0 \le \alpha_i \le C$$

Box constraints

Use SMO (Sequential Minimal Optimization) algorithm to solve

Non-Linear SVMs :-



$\nexists$ a "good" linear separator
$\exists$ a "good" quadratic separator

$\Rightarrow$ Next class.