non-separable solution

optimization problem:

$$\min_{\beta,\beta_0} \frac{1}{2}\|\beta\|^2$$

subject to $Y_i(\beta_0+\beta^T X_i) \geq 1-\xi_i \quad \forall i$

$\xi_i \geq 0, \quad \sum_{i=1}^{n} \xi_i \leq D$ (D is tuning param.)

Lagrange (primal) function:

$$L_P = \frac{1}{2}\|\beta\|^2 + \gamma \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i \Big(Y_i(\beta_0+\beta^T X_i) - (1-\xi_i)\Big) - \sum_{i=1}^{n} \mu_i \xi_i$$

tuning parameter

$$\frac{\partial}{\partial \beta} L_P = 0 \implies \beta = \sum_{i=1}^{n} \alpha_i Y_i X_i$$

$$\frac{\partial}{\partial \beta_0} L_P = 0 \implies 0 = \sum_{i=1}^{n} \alpha_i Y_i$$

$$\frac{\partial}{\partial \xi} L_P = 0 \implies \alpha_i = \gamma - \mu_i$$

substituting into L_P $\rightsquigarrow$ Wolfe dual

$$L_D(\alpha_1, \dots, \alpha_n) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \alpha_i \alpha_k Y_i Y_k X_i^T X_k$$

subject to $0 \leq \alpha_i \leq \gamma \quad \forall i$

$\sum_{i=1}^{n} \alpha_i Y_i = 0$

maximize $L_D(\alpha_1, \dots, \alpha_n) \rightsquigarrow \hat{\alpha}_1, \dots, \hat{\alpha}_n$

subject to $0 \leq \alpha_i \leq \gamma$, $\sum_{i=1}^{n} \alpha_i Y_i = 0$

solution:

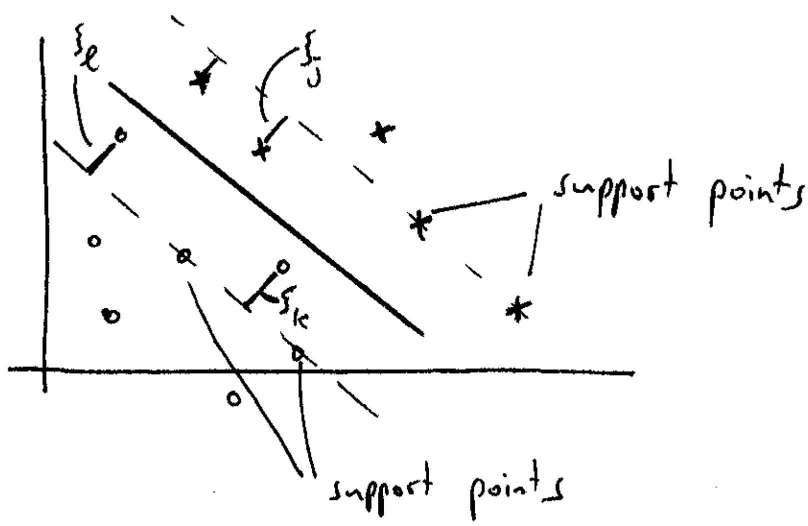
$$\hat{\beta} = \sum_{i=1}^{n} \hat{\alpha}_i Y_i X_i$$

support points: i with $\hat{\alpha}_i > 0$

$$\Longrightarrow Y_i \left(\hat{\beta}_0 + \hat{\beta}^T X_i \right) = 1 - \xi_i$$

$$\hat{\beta}_0 = \frac{1}{N} \sum_{i:\hat{\alpha}_i > 0} \left(Y_i (1 - \xi_i) - \hat{\beta}^T X_i \right)$$

$$N = \sum_{i=1}^{n} \mathbb{1}_{[\hat{\alpha}_i > 0]}$$



computation of $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}^T x$ involves inner products $\langle x, X_i \rangle$, $\langle X_i, X_k \rangle$