Programming assignment-2

EE15MH7P100001

Support Vector Machines & Duality

→ $\mathbb{R}^p$ [Binary classification]

SVM ⟶

standard SVM :-

→ Number of fields outside of just machine learning statistics.

standard SVM:-

higher price for misclassifications of one of the two-point clouds.

cost Sensitive SVM.

'n' data samples, each one taking the form (x_i, y_i) where $x_i \in R^p$ is a feature vector & $y_i \in \{-1, +1\}$ is a class

$$S_1 = \{ i \in (1, \dots n) ; y_i = +1 \}$$

$$S_2 = \{ i \in (1, 2, \dots n) : y_i = -1 \}$$

$[X \in R^{n \times p}]$

ⓐ Incorporate misclassification costs into the standard SVM formulation is to pose the following primal cost sensitive SVM optimization problem

$$\begin{cases} \min\limits_{\beta \in R^p,\ \beta_0 \in R,\ \xi \in R^n} \frac{1}{2}\|\beta\|_2^2 + C_1 \sum\limits_{i \in S_1} \xi_i + C_2 \sum\limits_{i \in S_2} \\ \xi_i \geq 0 ;\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i \end{cases}$$

$\beta \in R^p;\ \beta_0 \in R$ $\qquad i = 1, 2, \ldots n,$

$\xi = (\xi_1, \xi_2 \ldots \xi_n) \in R^n$ are our variable and C_1, C_2 are positive costs chosen by implementer.

① Does strong duality hold for Problem why or why not?

① ⓐ Strong duality

$$\min_{\beta \in R^p, \beta_0 \in R, \xi \in R^n} \frac{1}{2}\|\beta\|_2^2 + C_1 \sum_{i \in S_1} \xi_i + C_2 \sum_{i \in S_2} \xi_i$$

$$\xi_i \geq 0 \quad \{y_i [x_i^T \beta + \beta_0] \geq 1 - \xi_i\}$$

$$(i = 1, 2, \ldots n)$$

C_1 & C_2 are positive costs.

Strong dual :- $f^* = g^*$ [for primal & dual]

Slater condition :— If the primal is a convex problem. [f & $h_1, h_2 \ldots h_m$ are convex $I_1, I_2 \ldots I_r$ are affine] then there exist atleast one strictly feasible $x \in R^n$ meaning

$h_1(x) < 0 \ldots \quad I_1(x) = 0 \ldots \quad I_r(x) = 0$.

Then strong duality holds.

Here if we take w & b. & Let

$$a = \min_{1 \leq i \leq m} y_i (w^T x_i + b).$$

Define $\xi_i = (1 - a) \forall i$ then strictly feasible.

Defining $\xi_i = 1 - a \; \forall \; i$ the strict feasi-bility holds and slater's theorem implies strong duality.

[So. It ~~is~~ holds strong dual.]

② **KKT conditions for problem ①.**

$\alpha \in R^n$ for the dual variables

$\longrightarrow$ " $y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i$
$i = 1, 2, \ldots n$ " &

$u \in R^n$ for the dual variable for const

$\xi_i \geq 0 \quad i = 1, 2, \ldots n$

$$\min_{\beta \in R^p, \beta_0 \in R, \xi \in R^n} \frac{1}{2} \|\beta\|_2^2 + c_1 \sum_{i \in S_1} \xi_i + c_2 \sum_{i \in S_2} \xi_i$$

$$\text{s.t} \quad \xi_i \geq 0 ; \; y_i(x_i^T \beta + \beta_0) \geq 1 - \xi_i$$

$$(i = 1, 2, \ldots n)$$

Constraints formed :—

$g_i(\omega, b) = 1 - \xi_i - y_i(\omega^T x_i + b) \leq 0$

$h_i(\omega, b) = -\xi_i \leq 0 \quad (i = 1, \dots, n)$

Lagrangian :—

$L(\omega, b, \xi, \alpha, \mu) =$

$$\left[\frac{1}{2}\|\omega\|^2 + C_1 \sum_{i \in S_1} \xi_i + C_2 \sum_{i \in S_2} \xi_i - \sum_{i=1}^{m} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i \left[y_i(\omega^T x_i + b) + \xi_i - 1 \right]\right]$$

KKT conditions :— Stationarity :—

i) $\left\{ \frac{\partial L}{\partial \omega} \Rightarrow 0 \in \partial [L(\omega, b, \xi, \alpha, \mu)] \right.$ .. stationarity. $\}$

$\theta_D[\alpha, \mu] = \min_{\omega, b} L(\omega, b, \xi, \alpha, \mu)$

Derivate L with respect to ω, b, ξ to zero for ω:

$$\frac{\partial}{\partial \omega} L(\omega, b, \xi, \alpha, \mu) = \omega - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\Rightarrow \left[\omega = \sum_{i=1}^{n} \alpha_i y_i x_i\right]$$

for b

$$\left\{ \frac{\partial}{\partial b} L(\omega, b, \xi, \alpha, \mu) = 0 \right\}$$

$$\frac{\partial}{\partial b} L(\omega, b, \xi, \alpha, \mu) = 0 - \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\Rightarrow \left[\sum_{i=1}^{n} \alpha_i y_i = 0 \right]$$

for ξ_i :-

$$\frac{\partial}{\partial \xi_i} L(\omega, b, \xi, \alpha, \mu) = 0 \quad [i = 1, \ldots n]$$

$$C_1 - \alpha_i - \mu_i = 0 \quad \{ i \in S_1 \}$$

$$C_2 - \alpha_i - \mu_i = 0 \quad \{ i \in S_2 \}$$

(ii) Complementary Slackness Condition:-

$\{ (u_i h_i(x) = 0$ for all $i)$ — Basic KKT condition$\}$

$$\alpha_i \left(1 - \xi_i - y_i(\omega^T x_i + b)\right) = 0 \quad (i = 1, 2, \ldots, n)$$

$\mu_i \xi_i = 0 \qquad ; (i = 1, 2 \ldots n)$

(iii) Primary feasibility :-

$$1 - \xi_i - y_i (w^T x_i + b) \leq 0$$

$$-\xi_i \leq 0 \; ; \; (i = 1, 2, \ldots n)$$

(iv) Dual feasibility :-

$\alpha_i \geq 0 \qquad$ for all i

$\mu_i \geq 0 \qquad$ for all i.

(3) Derivation of Dual :-

$L(w, b, \xi, \alpha, \mu) =$

$$\left[\frac{1}{2}\|w\|^2 + C_1 \sum_{i \in S_1} \xi_i + C_2 \sum_{i \in S_2} \xi_i - \sum_{i=1}^{n} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i [y_i (w^T x_i + b) + \xi_i - 1]\right]$$

$$= \frac{1}{2}\left[\|w\|^2 + C_1 \sum_{i \in S_1} \xi_i + C_2 \sum_{i \in S_2} \xi_i - \sum_{i \in S_1} \mu_i \xi_i - \sum_{i \in S_2} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i (y_i (w^T x_i + b) + \xi_i - 1)\right]$$

from KKT condition

$$C_1 = \alpha_i + \mu_i \quad [i \in S_1]$$

$$C_2 = \alpha_i + \mu_i \quad [i \in S_2]$$

$$= \left[\frac{1}{2}\|\omega\|^2 + \sum_{i \in S_1} (\alpha_i + \mu_i)\xi_i + \sum_{i \in S_2} (\alpha_i + \mu_i)\xi_i - \sum_{i \in S_1} \mu_i \xi_i - \sum_{i \in S_2} \mu_i \xi_i - \sum_{i=1}^{n} \alpha_i [y_i [\omega^T x_i + b] + \xi_i - 1]\right]$$

By Canceling the terms.

$$= \frac{1}{2}\|\omega\|^2 + \sum_{i \in S_1} \alpha_i \xi_i + \sum_{i \in S_2} \alpha_i \xi_i - \sum_{i=1}^{n} \alpha_i [y_i(\omega^T x_i + b) + \xi_i - 1]$$

$$= \frac{1}{2}\|\omega\|^2 + \cancel{\sum_{i \in S_1} \alpha_i \xi_i} + \cancel{\sum_{i \in S_2} \alpha_i \xi_i} - \cancel{\sum_{i \in S_1} \alpha_i \xi_i} - \cancel{\sum_{i \in S_2} \alpha_i \xi_i} - \sum_{i=1}^{n} \alpha_i (y_i(\omega^T x_i + b) - 1)$$

$$= \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{n} \alpha_i [y_i [\omega^T x_i + b] - 1]$$

Scanned by CamScanner

by Substituting the KKT conditions

$$\left\{ \omega = \sum_{i=1}^{n} d_i y_i x_i \right\}$$

$$= \frac{1}{2} \sum_{i,j=1}^{n} d_i d_j y_i y_j x_i^T x_j - \sum_{i,j=1}^{n} d_i d_j y_i y_j x_i^T x_j$$

$$- \sum_{i=1}^{n} d_i y_i b + \sum_{i=1}^{n} d_i$$

$$= \sum_{i=1}^{n} d_i - \frac{1}{2} \sum_{i,j=1}^{n} d_i d_j y_i y_j x_i^T x_j$$

$$L = \max_{d} -\frac{1}{2} d \tilde{x} \tilde{x}^T d + 1^T d.$$

By the KKT condn derived

$$\left\{ \sum_{i=1}^{n} d_i y_i = 0 \right.$$

$$\left. y^T d = 0 \right\}$$

$$[\mu_i \geq 0] \Rightarrow C_1 - d_i - \mu_i = 0 \quad i \in S_1$$

$$C_2 - d_i - \mu_i = 0 \quad i \in S_2$$

$c_1 - \alpha_i \geq 0 \quad i \in S_1$

$c_1 \geq \alpha_i \quad i \in S_1$

$c_2 - \alpha_i \geq 0$

$c_2 \geq \alpha_i \quad i \in S_2$

$\alpha_i \leq c_2 \,; \quad \alpha_i \leq c_1$

$\alpha_i \geq 0$ [By dual problem].

$$\therefore \; w = \max_{\alpha} -\frac{1}{2} \alpha \tilde{X} \tilde{X}^T \alpha + 1^T \alpha .$$

$$\text{s.t} \quad y^T \alpha = 0$$

$$\left.\begin{aligned} 0 \leq \alpha_i \leq c_1 \quad \{ i \in S_1 \\ 0 \leq \alpha_i \leq c_2 \quad \{ i \in S_2 \end{aligned}\right\}$$

$$\left[\tilde{X} = \mathrm{diag}(y) x \right]$$

(7) Problem (1) & (2)

Both are quadratic programming as objective involves quadratic terms and linear constraints.

* quadratic Program.